

Whose right to genetic knowledge?

A new section in this issue of *Nature* highlights the growing demands of insurers for access to genetic information, the diversity of legislative responses, and the many scientific uncertainties. The history of genetics suggests a controversial way forward.

FEW implications of modern genetics generate as much heated debate as its consequences for medical and life insurance. The central issue is straightforward: information about an individual's genetic profile can provide important clues about his or her future state of health, and therefore has a direct bearing on actuarial calculations about that individual's health and life expectancy. But, as a series of articles in this week's issue demonstrates — serving also to inaugurate *Nature*'s new and occasional 'Briefing' section (pages 389–392) — how the implications should be tackled is less clear, and is the source of widespread disagreement. On the one hand, the insurance industry claims that access to genetic information is essential for the calculation of equitable premiums; on the other, the industry's critics warn of the undesirable effects that unregulated access could create.

The familial inheritance of certain disorders is already taken into account by insurance companies in determining whether to offer coverage to an individual, and, if so, on what terms. Meanwhile, gene-mapping and sequencing techniques, as well as the growing detailed knowledge of the relationship between genetic mutations and human disease, suggest that a rapid expansion is imminent in our ability to predict individual illness and mortality. Some late-onset illnesses (Huntington's disease is a well known example) can already be predicted with certainty from genetic mutations, whereas in some other diseases genetic information will at best deliver only an uncertain assessment of predisposition. But at both extremes of (un)certainly, such information is potentially equivalent in kind to other information already taken into account by insurers.

Nevertheless, public caution is widespread — partly because of lack of trust in insurers. The spread of HIV and AIDS has raised in an acute form the question of a company's right of access to the results of screening tests and how it should react to these results. And individuals already refuse to be screened for illnesses in case a positive result makes it impossible for them to obtain insurance at all. Such difficulties will increase rapidly as screening technologies develop.

It is easy, as some in the insurance industry are quick to point out, to exaggerate both our current abilities and the speed with which these are likely to grow. Much remains to be done in understanding not only how different mutations in individual genes may lead to a particular disease, but also how other diseases can result from an imbalance in the complex interaction of products of a number of different genes. Even more difficult to put into the quantitative terms required for actuarial calculations, is the impact, on both morbidity and mortality, of a combination of a genetic predisposition to a certain disease and the environmental factors that may either accelerate or delay its onset. For these reasons, perhaps, insurers are notably silent in response to enquiries about intended actuarial uses of genetic information.

In one sense, the interests of the industry and those it protects coincide: both seek to spread the burden of risk in an equitable fashion. But there is also an inherent conflict, in that both seek to maximize their own returns from the process. And there will inevitably be occasions, usually based on an imbalance of access to information, on which the efforts of the one to do so will

reduce the returns to the other.

Take, for example, the question of 'adverse selection'. This is the situation in which an individual finds through a genetic test that he or she has a high chance of developing a life-threatening disease within the near future, and takes out a large insurance policy on standard terms — those based on an assumption of a normal life expectancy. In this case, the individual concerned is likely to emerge a clear winner, while either the insurance company, or other insurance holders if the company decides to pass on its costs as higher premiums, will be the loser. Faced with such possibilities — which would be entirely legal under present rules, which do not require applicants to disclose details of a disease they are not currently suffering from — it is not surprising that the industry's nervousness is increasing.

There are other reasons for the insurers to fret — not least, issues of commercial competitiveness. For example, a company which offers cut-rate insurance to those who have taken voluntary tests that reveal a relatively 'healthy' genome not only stands to make a considerable profit, but will also raise the proportion of those disease-susceptibility genes among competitors who do not discriminate. Yet at the same time, the industry will win few friends if it does introduce widespread discrimination, because the relatively low number of individuals required to pay a higher than average rate for life insurance could well rise substantially.

Equally justified are concerns on behalf of those applying for both medical and life insurance. The presence of questions about genetic testing on insurance application forms could deter some individuals from volunteering for screening tests that would otherwise be desirable. At the same time, there are ethical dilemmas for those who may prefer not to know the results of such tests (for example, the prospect of developing Alzheimer's disease).

As the years pass, the development of genetic tests and therapies will lead to a growth in confidence in the handling of such information. In the meantime, regulation is required. Here the successful experience of handling the safety aspects of genetic engineering provides a useful model; a short, sharp moratorium on use of genetic information by insurers combined with fierce public debate, followed by initial regulations that are deliberately over-cautious, and a subsequent period of liberalization as more becomes known of both the real and hypothetical dangers posed. Despite short-term problems and controversies, this is the best way in the longer term to develop public confidence in understanding both the tests and the way in which they are applied actuarially, and in the ability of regulators to curb abuse. But in the end, both insurer and policy-holder will have to share information equitably — there is no long-term alternative to consensual discrimination.

These issues loom large, but need also to be seen in the context of other far-reaching implications of current research, not least the future need (and desirability) for people and societies to adjust more generally to the predictabilities that will arise from genetic knowledge. Treatments of illness will also advance, but the scale of the necessary rethinking about our lives will make today's debates about insurance seem relatively trivial. □



Indian researchers press for stricter rules to regulate 'gene-hunting'

New Delhi. The Indian government is coming under pressure from some of its own scientists to tighten the rules on the export of human DNA and blood samples. This follows growing evidence of the interest of foreign research groups in using India as a fruitful 'hunting-ground' for disease genes and their mutations.

Much of the pressure is coming from Indian researchers concerned that some valuable genetic information appears to be being taken from the country 'illegally' under the guise of collaborative research in human genetics.

Since 1992, the export of biological tissues has required specific clearance from the Indian Council of Medical Research (ICMR). But the council lacks sufficient powers to enforce the rule, and its officials say that they therefore feel unable to do anything about samples leaving India without their knowledge (see box below).

With its huge population of diverse cultures and significant in-breeding, India is a natural setting to look for rare genetic mutations. The Onge tribes of India, for example, have a small Y chromosome and low sperm

count, while a community of 700 families in southern India suffers from a combination of osteoarthritis and dwarfism.

"Name any genetic disorder and we have the mutations," says Samir Brahmachari, professor of molecular biophysics at the Indian Institute of Science in Bangalore. "If you are looking for twins, we have plenty."

India also provides scope for "differential genome comparison". In West Bengal, for instance, where cholera is endemic, a large group of individuals appears to be immune to the disease. The gene responsible, which may eventually lead to a valuable commercial product, is now being sought at a laboratory in Calcutta using sophisticated equipment donated without strings by the government of Japan.

In addition, Indian geneticists point out that, with family structures crumbling in the West, the data about large families required for linkage studies are available only in India because China — the other populous country — is not easily accessible.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Family affairs: the incidence of large families and in-breeding makes India attractive to geneticists.

But the growing interest of foreign research groups is prompting the Department of Biotechnology (DBT) to tighten rules on the export of genetic data, and to draw up guidelines for international collaboration before the launch of the Human Genome Diversity Project (HGDP), a global effort to study genetic variations in *Homo sapiens* (see *Nature* 377, 372; 1995).

The latter project is planning to study blood samples taken from indigenous populations worldwide, including 23 ethnic ►

Scientists challenged over 'unauthorized' export of data

New Delhi & Washington. Researchers from the US National Institutes of Health (NIH) have been accused of violating an Indian government regulation forbidding the export of biological material without specific permission from the Indian Council of Medical Research (ICMR).

According to a spokesman for the ICMR, the NIH's National Eye Institute (NEI), which is collecting DNA and blood samples from patients at major private eye hospitals in India, has not obtained the authorization required to take samples out of the country, and is therefore acting "illegally".

But Carl Kupfer, the director of the NEI, while confirming that the institute is receiving blood samples from India, points out that the research protocols have been cleared by review boards in both India and the United States. He adds that neither the NEI, nor the Indian hospital with which it is collaborating, is aware of any requirement to clear the arrangement with the ICMR.

The US scientists are searching for the gene or genes that cause retinitis

pigmentosa (RP), otherwise known as night blindness. India is an ideal social setting for mapping the location of the gene because of its large families — several members of which can suffer from RP — and the high incidence of intermarriage.

The NEI is already collecting samples from L. V. Prasad (LVP) Eye Hospital in Hyderabad, and is engaged in discussions to collect similar samples from three others, the Sankar Nethralaya in Madras, Little Flower Medical Centre in Angamally in Kerala, and an eye hospital in Amritsar. The last of these has a large database with details of more than 20,000 patients suffering from RP and congenital cataract.

According to Subadra Jalali, a surgeon at the LVP hospital, blood samples of 80 members belonging to six RP-affected families have already been sent to the NIH, and the hospital has agreed to provide such data on 20 families over a period of three years. In return, says Jalali, the NIH will provide the hospital with \$180,000 to cover payments to patients who give blood

"voluntarily".

Government officials claim that NIH is not the only foreign body which, they say, is exporting without authorization Indian blood samples. Researchers at the Duke University Medical Centre in Durham, North Carolina, for example, are close to identifying the mutation responsible for juvenile myoclonic epilepsy (JME) which accounts for ten per cent of all epilepsies.

The US scientists were fortunate to find a clinician in a Delhi hospital who provided them with blood samples from what is said to be the world's largest family with JME. Nine of its 30 members are affected.

In response to the charge that US researchers are bypassing the views of leading Indian geneticists and ignoring local export regulations, Kupfer points out that the NEI project is being carried out in collaboration with J. S. Murty of the University of Osmania at Hyderabad, a past-president of the Indian Society of Human Genetics, and that it includes a strong training component.

K. S. Jayaraman and Colin Macilwain

► groups from the Indian subcontinent. But several members of DBT's own task force on human genetics warn that, in the absence of strict regulations, the country's unique human genetic diversity might be exploited by foreigners without India receiving any tangible benefits.

Last month, the Indian Society of Human Genetics (ISHG), meeting in Calcutta, drew up guidelines calling for a ban on transport of "whole blood, cell-lines, DNA, skeletal, and fossil material" without a formal agreement between collaborating parties.

According to the society, any such agreement should be approved by the appropriate government agencies and should clearly specify "the objectives of the project and the anticipated scientific, material, and economic benefits and the manner that they are to be shared both at present and in future".

Partha Mazumdar of the Indian Statistical Institute in Calcutta, a member of the national committee on HGDP, says that even if DNA is shared for research purposes, "subsequent transfer must clearly be stopped, as the end-use of the specimen cannot be overseen".

But some clinicians disagree, arguing that the results of the project will eventually benefit India's population. "We have clinical materials, [Western countries] have the technology, and we can work together with a good understanding," says Satish Jain, a neurologist at the All India Institute of Medical Sciences in New Delhi.

Jain, who is sending blood samples taken from epileptic patients to researchers in the United States, Canada and Australia without ICMR clearance, says he believes that government regulations will kill collaborative projects with foreign research teams.

In general, Indian scientists agree that collaboration with Western research groups is essential, as there are very few laboratories in India sufficiently well equipped to study genetic disorders in the population. But there is concern that the results of collaboration may prove one-sided.

"It must be on an equal footing, and on a formal basis," says Brahmachari, a member of the DBT task force.

Indian scientists say that the global map of genome diversity will be incomplete without mutation data from regions such as India and that, while the government should not obstruct sharing of the data, its potential commercial value (and ownership) should be protected by appropriate policies.

The DBT is unlikely to announce a policy decision until a new secretary is appointed to succeed Chittaranjan Bhatia, its former head, who retired in December. Before his departure, Bhatia had described the department's mission as being to "convert India's genetic wealth into economic wealth".

To this end, the DBT has already begun its own programme on human genome diversity research. Four regions have been chosen for studies to identify genes responsible for diseases.

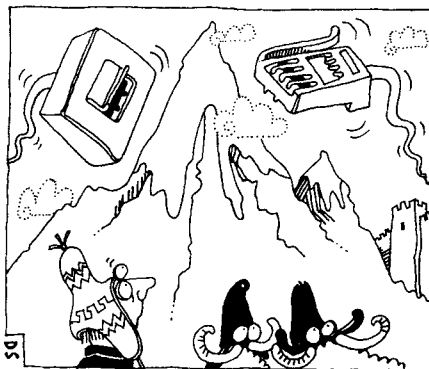
K. S. Jayaraman

Internet struggles to connect around the Asia-Pacific rim

Hong Kong & Tokyo. Both the private sector and government agencies in the Asia-Pacific region are moving rapidly to increase the capacity of Internet connections between countries on the Western Pacific rim. But many observers fear that their efforts may be premature, given the relatively undeveloped level of computer networks in many countries of the region.

Last month, Asia Internet Holding Company Ltd, a new joint venture set up between companies in Japan, Hong Kong and Singapore, took out a lease on a high-capacity Internet link between Japan and Hong Kong, and announced that it plans to lease further lines, in order to create an 'Asian backbone' running between Japan, Hong Kong and Singapore.

In addition, the Japanese government is floating the ambitious ideas of building a gigabit capacity network linking researchers of the Asia-Pacific region, following discus-



sions during a meeting of science ministers of the Asia Pacific Economic Cooperation (APEC) forum in Beijing last year.

Japan, South Korea, Taiwan, Hong Kong, Singapore and Australia are relatively well connected to the Internet through individual high-capacity links to the United States. But the capacity of connections between Western Pacific rim nations is limited, and most intraregional traffic is therefore routed via the West coast of the United States — one reason why some of the trans-Pacific links, particularly those between Japan and the United States, cannot meet demand.

Asia Internet Holding Company was set up last October by Internet Initiative Japan (IIJ), one of Japan's new commercial providers of Internet access, Sumitomo Corporation, a large Japanese general trading company, Pacific Internet of Singapore-owned Sembawang Corporation, a state-backed shipbuilding company, and Hong Kong Supernet Ltd, a commercial access provider set up by Hong Kong University of Science and Technology (HKUST).

The company has leased a line capable of

transmitting 768 kilobits per second between Japan and Hong Kong. This will be upgraded to T1 capacity (1.2 megabits per second) this month, and the consortium also plans to establish another high-capacity link between Japan and Singapore in the near future, according to Eugene Wong of HKUST, who set up Supernet.

The joint venture has yet to be granted operating licences from all the governments concerned. But Wong says customers are already queuing up to use the service, and that the new company expects business to flourish with the rapid growth of Asia's economies.

Other Internet providers in Japan remain sceptical about whether the need yet exists for such links, and are concentrating on expanding trans-Pacific capacity. But Wong says he is convinced that the joint venture will have a "catalytic" effect on the development of the Internet in Asia, and that the high capacity lines will provide a "natural electronic backbone".

The Japanese government is also keen to develop computer networks in the Asia-Pacific region. Japan's Science and Technology Agency (STA) will host a symposium next month in Tsukuba science city to discuss the prospects for a gigabit network linking the Asia-Pacific region. But such ambitious ideas represent a major step from the present level of computer networks in many Asian countries.

China, for example, installed its first 64-kilobit link to the outside world only in 1994, and the Chinese government may well hold back further development of the country's computer links by trying to restrict the commercial development of the Internet, just at a time when the private sector is becoming the driving force behind expansion of the network elsewhere in the world.

Last week, China's State Council announced plans for new regulations to strengthen government control of online links to the outside world. This follows a decree passed by the Communist party and the state council at the end of December acknowledging that the Internet is important for the economy and science, but warning that it threatens to usher in pornography and other "harmful materials".

Another reflection of the primitive state of Internet connections in Asia is the fact that the STA is planning to distribute information about Japan's government research institutes to countries in the Asia-Pacific region by CD-ROM rather than the Internet. The sum of ¥130 million (US\$1.2 million) has been set aside for this purpose in the agency's budget for 1996.

David Swinbanks

Bleak prospects for physics research council

London. Britain's Particle Physics and Astronomy Research Council (PPARC), set up two years ago as the main channel of funding for basic research in these two disciplines, has warned that a lack of new programmes resulting from pressures on government spending means that Britain could be in danger of dropping out of the top league in science.

In evidence presented to an inquiry on the council being carried out by the House of Commons Select Committee on Science and Technology, PPARC points out that, because of these pressures, it has been able to launch virtually no major new scientific projects since coming into existence. The main exception is its commitment to participate in the construction of the Large Hadron Collider (LHC) at the European Laboratory for Particle Physics (CERN).

In astronomy, for example, it has had to turn down an invitation to contribute a key sensor to a Japanese space-based astronomy mission, and to limit severely its participation in the joint construction with Germany of a gravitational wave observatory. It has also been unable to find additional money to develop experiments with an underground detector for particles that may account for the missing 'dark matter' in the Universe.

"There has been a form of planning blight hanging over all our programmes," says Ken Pounds, PPARC's chief executive officer and professor of radioastronomy at the University of Leicester. "The analogy that one colleague has drawn is that until recently we have been driving toward the cliff edge, now we are driving along the cliff edge."

PPARC does not blame its problems entirely on the government. One difficulty has been the increasing costs of its subscription to CERN, the European Laboratory for Particle Physics, due to the fall in value of the pound against the Swiss franc (see *Nature* 379, 286; 1996). Another has been the extent to which its hands have been tied by various commitments — such as that towards the Gemini telescope — made in the final days of its predecessor, the Science and Engineering Research Council.

At the same time, however, PPARC quotes two additional factors in its evidence to the select committee. It acknowledges that its plight is partly due to decisions taken to enhance its programmes in education and training and in building links to industry, as well as to "the failure, so far, of any additional bids to government for new funds" (with the exception of ring-fenced allocation for CERN exchange rate variations).

PPARC is unlikely to gain sympathy from all other sectors of the research community, particularly from those who feel that Britain still spends too much on 'big science'. Giving oral evidence to the select committee last

week, Sir John Cadogan, the director general of research councils, claimed that, on the basis of a review he had carried out last year, "the majority of the science base does not believe that particle physics deserves its present level of funding".

But PPARC's response is to point out that, in relative terms, its share of total science funding has dropped significantly in recent years. Figures presented to the committee show that, despite an increase of about 53 per cent in real terms over the past 20 years, Britain's spending on science, total expenditure on particle physics and astronomy has fallen by 32 per cent during this period — and the rising costs of international subscriptions means that domestic spending has dropped even further, by 53 per cent.

At the same time, the council claims it has evidence to show that "UK research productivity in most PPARC areas exceeds that of our major competitors". The council's main concern is that restrictions in the domestic budget mean that it is able to participate in international programmes at only about half the level that it should be

doing on the basis of its contribution to the costs of major facilities such as the LHC.

It also expresses concern at the amount of management time and effort being required from a reduced number of headquarters staff to participate in a range of initiatives launched by the Office of Science and Technology (OST), including management reviews, the Technology Foresight Programme and the so-called ROPAs awards scheme. Ironically, it only received two applications for such awards, designed to stimulate academic/industry collaboration, in 1995, neither of which was judged by outside assessors to meet the criteria laid down for such awards by the OST.

"PPARC is currently maintaining a strong science programme in particle physics, planetary science and astronomy, but these opportunities are based largely on earlier investments," the research council warns. "Severe funding pressures on PPARC's domestic budget is now limiting future investment which will increasingly threaten the health of the research communities into the next century."

David Dickson

Leviathan telescope set to rise again

London. A giant 150-year-old telescope, believed to be the world's largest when it was built in 1845, is to be rebuilt at its original site in the grounds of a seventeenth century castle in Ireland, the proposed location of a new Historic Science Centre.

The 72-inch *Leviathan* (right), which had a focal length of 54 feet, was built and erected by William Parsons, third Earl of Rosse and one-time president of the Royal Society, at the family estate of Birr Castle in County Offaly.

The *Leviathan* took 20 years to complete and cost £30,000. It weighed more than 150 tonnes and needed four men to operate it. Parsons, who became a member of parliament at the age of 22, constructed the 4-tonne mirror himself. The rest of the telescope was built with the help of his servants and other local people.

A grandson dismantled the telescope in 1908, when it became old and rickety. The IR£3-million (US\$4.67-million) reconstruction is being supervised by the present Earl of Rosse. The mirror will be made by the optical

department at University College, London, and the telescope is expected to be completed by March 1997.

Despite Ireland's often unclear weather, Parsons managed to make a

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

number of important discoveries. He is believed to have been the first astronomer to make out the spiral shapes of cloudy objects in the night sky that were later recognized as independent galaxies such as our own Milky Way.

In 1848, he studied an irregular foggy patch of stars which he thought resembled a crab. The name of the patch, the crab nebula, survives to this day.

Ehsan Masood

Poor data hampers East–West cooperation

Munich. Scientific cooperation between Eastern and Western Europe is being hindered by a lack of reliable data about science and technology activities in Central and Eastern Europe, according to the parliamentary assembly of the Council of Europe, which last week passed a resolution calling for the establishment of a system to collect such statistics.

The assembly also called for cooperation programmes to give priority to increasing support for projects facilitating the conversion of military research projects to civil goals, pointing out that efforts to do this have so far been notably unsuccessful.

Such resolutions are not binding on the Council's 38 member states. But a spokesman for the Council's secretariat says that they are intended to focus the attention of the West on the serious problems facing many Central and Eastern European countries as they attempt to reform their scientific efforts after the fall of communism.

The assembly's decisions follow a report from the Council's Committee on Science and Technology pointing to the disappointingly slow transformation of Eastern European research structures into the standard Western model, which it describes as being based on openness, competitiveness and the independent assessment of the quality of research activities.

According to the report, which acknowledges considerable differences in the rate of reform in individual countries, resistance to reform has been fuelled by, among other factors, the frequent unwillingness of academies of sciences to give up the many privileges they enjoyed in communist times, and by a 'culture of secrecy' which resists pressures for public accountability.

Central and Eastern Europe still has

need of the West to help the reform process, it says. But such help should be given in terms of concrete investment — that is, financial and material resources — rather than merely advice. The committee claims that Central and Eastern European countries are starting to feel patronized by the well-meaning gestures of Western organizations who over the past few years have sent numerous delegations to give advisory seminars or pass judgement on quality of research but have brought little in terms of money or scientific equipment.

The resolution passed last week also calls for a clearer focus for cooperation agreements, and stresses the urgency of projects to 'convert' military research.

In addition, research agreements have been hindered by a continued deterioration

in the research infrastructure in Central and Eastern Europe, as well as by severe bureaucratic hurdles. The report also hints at problems caused by uncertainties over where Western money goes once it arrives in the East, due to inadequate public records. This would be helped by the creation of a system, proposed in the resolution, of formal statistical accounting.

But there is one hopeful sign in the report. New estimates of the size of the East-to-West brain drain suggests that the problem may have been exaggerated in the past. After the major wave of emigration of scientists in the early 1990s, the situation has now stabilized, it says, and current loss to the West is believed to total less than 10 per cent of the total research and development capacity in each country. **Alison Abbott**

Materials advice for German research

Munich. Germany is to set up a national advisory group for materials sciences in the light of criticism from its Science Council, the Wissenschaftsrat, of the lack of communication both among research institutes and between research institutes and industry. The report blames this lack of coordination for Germany's failure to remain internationally competitive in the development of industrially relevant new materials.

Franz Mayinger, an engineer from the Technical University of Munich who headed the committee that prepared the report, says that the new advisory group, which is expected to be financed initially by the federal research ministry (BMBF) before finding a home within industry, will seek to encourage interdisciplinary coordination and to promote the more efficient transfer of tech-

nology between laboratories and industry. Its members will include both scientists and local and federal government officials.

The expert committee spent the past 18 months reviewing the work of 34 of the country's publicly funded non-university research institutes that undertake materials research, and concluded that standards varied widely. The large Max Planck Institute for Polymer Research in Mainz, for example, was singled out for praise, while the Fraunhofer Institute for Solar Energy Systems in Freiburg was heavily criticized for the low quality of its work.

There was also heavy criticism of the materials science programme within the struggling national research centre at Karlsruhe, recently renamed the Forschungszentrum Karlsruhe. The centre was established in the 1950s as a nuclear research centre, and has since been forced to turn to new research directions, but the commission said that its material science efforts were too diverse in their aims and thus unproductive. This judgement puts more pressure on the Karlsruhe centre which, in the eyes of many politicians, has so far been unsuccessful in making a smooth transition from nuclear to non-nuclear research.

Unlike previous Wissenschaftsrat reports, this one does not recommend the closure or contraction of any research institutes. But the report does make the usual plea for funds. The long-term nature of new materials development, which can take up to 20 years, requires more generous resource commitments from the BMBF, which usually restricts project funding to five years, says Mayinger. He adds that the advisory group will also address the industry's apparent lack of interest in funding new materials research because of pressure arising from Germany's current economic recession. **A. A.**

Russia approves gene therapy research grants

Moscow. The Russian Human Genome Project (RHGP), recently the subject of international scientific concern (see *Nature* 374, 580; 1995), has overcome its organizational difficulties and made an important step forward with the approval of grants to a number of research projects aimed at the development of gene therapy.

The grants cover various fields. Four deal with the development of approaches to the introduction of dystrophin gene and/or its 'minigenes' into muscles as a possible treatment for Duchenne and Becker myodystrophies. Two are devoted to the use of antisense genetic constructions as potential tools for anticancer therapy, and one to transferring the human apolipoprotein gene into liver cells both *in vitro* and *in vivo* as a possible approach to atherosclerosis gene therapy.

Special attention is given to the

development and application of 'ballistic' techniques for inserting genes into their targets. "This is a field in which Russian science has a clearly defined priority," says Lev Kisselev from the Engelhardt Institute of Molecular Biology in Moscow, who was recently appointed head of the RHGP. The RHGP council has also decided to set up a committee to decide on the conditions under which human gene therapy should be carried out in Russia.

The success of the gene therapy projects will depend heavily on the amount of money that the Russian government, including the Ministry of Science and Technological Policy, agrees to allocate for the RHGP as a whole, and to the human gene therapy programme in particular. Until now, funding for the RHGP has been very modest, totalling only about US\$500 000 for the whole 1995. **Carl Levitin**

IMAGE
UNIMAGE
FOR A COPY
FOR A COPY
REASONS
REASONS

Patrick Kovanick/AFP

US court challenge may limit use of data in DNA banks

Lucas: health secretary asked him to stand.

Cancer charity installs former critic as head

Paris. The financial scandal at France's biggest medical charity, L'Association pour la Recherche contre le Cancer (ARC), took an ironic twist last week when the executive board voted to replace Jacques Crozemarie, its chairman, with the former head of the government agency that had fought for more than a decade to expose its financial irregularities.

The new chairman is Michel Lucas, previously head of the Inspection Générale des Affaires Sociales (IGAS). In 1990, Crozemarie blocked an investigation of ARC by IGAS by winning a court ruling that the agency — which polices public bodies — had no right to investigate private charities. But the preliminary findings of the investigation were leaked to the press, and prompted an investigation by the national audit commission.

The audit commission has now confirmed the main conclusions of the IGAS investigation, namely that only one-quarter of the charity's spending went on research (see *Nature* 379, 103; 1995). It also found evidence of fraud in the activities of companies associated with the charity, which are being investigated by the public prosecutor.

Lucas agreed to his nomination only a few hours before an extraordinary meeting of the charity's executive board took place last week, and following a personal appeal by Hervé Gaymard, the French secretary of state for health.

In a secret ballot, 17 members of the board voted for Lucas, and 11 against. At the same meeting, 18 voted to accept the conclusions of the auditor's report, and 10 voted against.

Lucas's first action was to freeze ARC's accounts and to require that all payments received his personal approval. He also promised that the funds allocated to research will increase "rapidly" to one-half of the charity's spending, to clarify the role of ARC's various decision-making bodies, and to establish the "transparency" needed to reestablish the confidence of donors. Lucas also said that ARC would add its name as a plaintiff in the fraud investigations.

Declan Butler

San Francisco. Limitations on the government use of DNA samples stored in genetic databanks are expected to emerge from a court challenge to the collection of such samples by US military authorities.

The case has recently entered an appeals court mediation process in San Francisco. Paul Billings, acting associate professor of medicine at Stanford University, suggests that the outcome could mean that the largest DNA bank in the world will have changed from being entirely unrestricted, to recognizing the dangers associated with the storage of genetic information.

The case began in January last year, when John Mayfield and Joseph Vlacovsky, both members of the US Marine Corps stationed in Hawaii, were asked for a blood sample and a swab taken from the inside of their cheeks for the Department of Defense's DNA Specimen Repository. The military has been collecting such samples since 1992 in order to help identify the remains of those who may be killed in action.

In what appears to be the first challenge to an employer's mandatory DNA collection programme, the two individuals refused to provide the required samples, and were court-martialled. Their military case is now on appeal, but meanwhile the two filed a civil suit against the government. Both sides agreed in December to seek mediation through the US Court of Appeals for the Ninth Circuit in San Francisco.

Mayfield and Vlacovsky are objecting to the enforcement of a programme that requires the collection of bodily tissues without their consent, according to Eric Seitz, a civil rights attorney in Honolulu, Hawaii, who is representing the two men. Seitz says that they are particularly concerned that military authorities are not prohibited from using DNA samples for purposes other than the identification of remains.

According to documents that have been lodged with the court, the government set up the database following problems encountered with the identification of individuals who were killed during the 1991 Gulf War. DNA analysis of remains proved more useful than fingerprint or dental records. But it required intrusion into the privacy of family members in order to collect blood samples, and in at least one case raised a question of paternity.

According to Lieutenant Colonel Victor Weedn, the programme manager of the registry, a DNA bank of all those who are in active service is the logical solution, and so far, the bank has collected more than a million samples.

The databank stores the dried blood samples on vacuum-sealed cards placed

inside a metal-foil pouch, and the oral swabs in a vial of isopropanol. Both are kept in a secure building, and are available only for limited uses, with access restricted to the resulting information.

The samples can at present be made available for other DNA profile analyses through a formal request to the Assistant Secretary of Defense for Health Affairs, although Weedn says that he knows of no such request having been made or approved so far.

Seitz says that the two men are not trying to shut down the databank; they merely want the collection of samples to be voluntary, and for procedures to ensure that the material can be used only to identify remains. They also believe that samples should be returned or destroyed when an individual has left the military.

According to Weedn, however, the military plans to keep samples from individuals for 75 years before destroying them. This, he says, is because it may take many years to identify remains; in fact, some are still being identified from the Second World War, Korea and Vietnam. He also claims that it would be too time-consuming and costly to return the material to the 200,000 or so members discharged from the military each year.

But Seitz says that his clients remain worried, and would be far less concerned if the whole process were to be based on a process that required the informed consent of participants. He cites various historical examples in which genetics knowledge has been mishandled by the military, such as its use during the mid-1970s to exclude applicants with sickle-cell trait from the Air Force Academy and to disqualify them from aviation and flight crew training in any military department.

Seitz points out that genetic data can be used to analyse predisposition to illnesses, and may be correlated with behavioural characteristics and sexual orientation. "It's too dangerous and explosive of an area for the government to engage in without limitations," he says.

But in documents that have been filed with the San Francisco court, the Marine Corps claims that the plaintiffs have misinterpreted the scope and operation of the DNA bank. It also asserts that all service members have implied their consent to such a programme through their voluntary enlistment, citing previous civil court rulings that military personnel are entitled to less privacy than other citizens, and referring to courts that have upheld the mandatory collection of DNA samples from convicted criminals.

Sally Lehrman

Embryo research barred from federal funds

Washington. In a move that seems destined to cast a new chill over research on human embryos in the United States, President Bill Clinton last week signed into law a ban on US government funding of such research, a concession that formed part of a package aimed at averting a further shutdown of parts of the government.

The ban was inserted by Republicans in a temporary spending measure passed by the House of Representatives on Thursday (25 January), and by the Senate on the following day. The bill, which was signed by Clinton the same day, funds those departments and agencies that still lack permanent 1996 funding until 15 March.

The ban on funding for embryo research,

however, which had been dropped from the bill during earlier negotiations, will apply to the end of the 1996 fiscal year on 30 September. It is thought to have been reinserted by Republican negotiators in order to pacify those in their ranks who had softened their position on a separate, anti-abortion, provision. "It was a sweetener for the pro-life crowd," admits one Republican staff member.

In practice, the ban will not halt any current research, as none is being funded by the government. But it reduces the chances that the National Institutes of Health (NIH) will try to initiate embryo research in the next fiscal year, despite the recommendations of an expert panel which concluded in

1994 that, within stringent constraints, research on human embryos is acceptable before the primitive streak appears, at 14 days (see *Nature* 376, 288; 1995).

Since then, what appears to have been fear of political fallout in the Republican-dominated Congress has kept Harold Varmus, the director of the NIH, from enacting the panel's recommendations and thus opening the door to NIH-sponsored embryo research.

The White House argues that Clinton's signing of the budget bill would thus make no difference to the NIH's research plans. "This provision would have no effect on any research currently being funded by NIH," said one official. In 1994 Clinton banned federal funding for research on embryos fertilized exclusively for research purposes. He issued the ban on the same day that the NIH panel concluded that such research was acceptable in limited circumstances.

The new, broader ban has pleased anti-abortion Republican legislators. "It will serve a very good purpose in our society because it will honour the sanctity of life," said Jay Dickey (Republican, Arkansas), one of the original authors of the ban.

This was accepted by the House last summer, as part of the bill funding the Departments of Labor, of Health and Human Services and of Education, although a companion bill stalled in the Senate, giving rise to the current need for a temporary funding measure.

But many researchers are dismayed, in particular over the fact that the administration has made a significant concession to conservative Republicans on the issue. "If this persists, there is a danger that some of the important work in the field of embryo research will be delayed. It will impact ultimately on patient care," says Zev Rosenwaks, director of the Center for Reproductive Medicine and Infertility at Cornell University Medical College in Ithaca, New York.

Similarly, Roger Pedersen, a professor of obstetrics, gynaecology and reproductive sciences at the University of California, San Francisco, claims that the ban is "a great loss for millions of Americans suffering from infertility" and that federal support for human embryo research "should not be made a sacrificial lamb for the sake of this political deal".

Although some embryo research, which aims to improve infertility treatment by studying early embryonic development, is still being carried out in the United States, the current ban means that it can be done only with private money. Pedersen estimates that about 300 US centres offer *in vitro* fertilization procedures, but only about two dozen of these carry out significant research on human embryos.

Meredith Wadman

Shot in the arm for medical policy body

Washington. A US health research agency whose budget faced the threat of a 60 percent cut by Republicans in the House of Representatives has become an unwitting, if qualified, beneficiary of Washington's current budget impasse.

The proposed 1996 budget for the Agency for Health Care Policy and Research (AHCPR), which carries out research aimed at improving the quality and efficiency of health-care in the United States, was reduced last August by Republicans in the House of Representatives to \$66 million from its level of \$159.3 million in 1995.

But under several temporary spending measures that have served since October to keep unfunded parts of the US government functioning, AHCPR is being funded at a level equivalent to \$119.5 million — 75 percent of its 1995 budget. The most recent stop-gap measure, signed by President Clinton last week, continues this funding level through 15 March, when nearly half of the 1996 fiscal year will be over.

This turn of events has angered the Republicans responsible for proposing the original funding reduction, who claim that the agency, best known for its development of clinical practice guidelines, duplicates the work of private medical associations and researchers.

"I'm going to do everything I can to get them off that [\$119.5 million] number in the next continuing resolution," says Sam Johnson (Republican, Texas), author of the 1995 amendment that cut the agency's funding to \$66 million.

Johnson argues that the agency is trying to micro-manage US doctors and hospitals from Washington. "No one has yet told me what they do for America that's good," he says.

The agency's work has also angered some medical pressure groups. The Republicans had been supported in their attempts to cut the budget by the Center for Patient Advocacy, in which a number of spinal surgeons are active. This group disagrees with AHCPR guidelines on lower back pain which claim that surgery is usually unnecessary if the pain lasts less than three months (see *Nature* 377, 379; 1995).

AHCPR research also angered cataract surgeons three years ago by concluding that surgery was not always the best treatment for cataracts.

But officials from the agency defended its mission vigorously at a meeting of its national advisory council last week. They argue that its work would not and could not be absorbed by the private sector, and that the savings to health-care resulting from the research of the AHCPR significantly outweigh the dollars that have been spent on funding it.

Officials also complain that the current funding uncertainty is hurting the agency, which has seen applications for its research grants fall by nearly half since its fiscal crisis began last year. The six-year-old agency was left without agreement on an annual budget following the failure of a Senate bill that would have funded it more generously than the House, at \$127.3 million, in 1996.

Clifton Gaus, AHCPR's administrator, describes it as "ironic" that the agency's future has been left hanging "at a time when our work is probably more important than it's ever been". He points out that the rapid shift in the United States toward managed health-care has highlighted the importance of gathering data on the cost efficiency and clinical effectiveness of medical procedures. **M. W.**

Panel backs closure of US fusion machine

Washington. The largest fusion machine in the United States should be closed down next year, and the national fusion research programme refocused on basic plasma physics and materials science, according to a key advisory committee to the Department of Energy (DoE).

These conclusions, which would mean the closure of the Tokamak Fusion Test Reactor (TFTR) at Princeton, New Jersey, were formally presented by the DoE's Fusion Energy Advisory Committee (FEAC) to Martha Krebs, director of its Office of Energy Research, on 27 January. They were immediately accepted by Krebs and, as the likely basis of a scaled-back US fusion research programme, are due to be considered by Congress in a series of hearings planned to take place in Washington next month.

The US fusion budget was cut back by one-third in the current financial year to \$244 million, when the Tokamak Physics Experiment (TPX), a major new experiment planned for the Princeton Plasma Physics Laboratory (PPPL), was cancelled by Congress (see *Nature* 377, 567; 1995).

For the next financial year, which starts in October, the Clinton administration is expected to propose a budget of about \$250 million, or slightly higher. Congress, having singled out fusion for attack last year, may accept the administration's budget proposal this time — provided, according to Congressional staff members, that the fusion community agrees on how it should be spent.

FEAC's proposal would allow the United States' other two large tokamak facilities, namely General Atomics' DIII-D machine at San Diego, and the C-Mod Alcator tokamak at the Massachusetts Institute of Technology, to operate at high levels of utilization.

It would also allow the US to continue participating in the design phase of the International Thermonuclear Experimental Reactor (ITER) up to the completion of this phase in 1998, and for a modest expansion in support for small university teams working on plasma physics and materials science.

But according to FEAC, if the fusion budget is set at \$250 million, TFTR could only operate briefly during 1997 before closing down for good, a year earlier than planned. The advisory committee adds that PPPL, of which TFTR is the centrepiece, is "a critical national resource for the fusion programme" and must be retained. "We can't imagine the programme without it," says Mike Knotek of the Batelle Pacific Northwest Laboratories, chair of the panel which drafted the committee's report.

Many fusion scientists, however, doubt that PPPL can keep its leadership role after TFTR shuts down. The laboratory is there-

fore planning to continue to press for the retention of TFTR, at least until 1998. It argues that there are valuable experiments, especially concerning the sheared flow of plasma and using deuterium-tritium plasmas, which no other machine in the world can perform.

Dale Meade, deputy director of the laboratory, says that TFTR could continue to operate within a \$250-million budget, and that the only purpose of its early closure would be to appease Congress's desire for a sacrificial lamb. "If you want to maximize



Threatened: the Tokamak Fusion Test Reactor at Princeton would close under the plan.

the best science, you can run it with the budget at \$250 million," says Meade. "If you want to maximize the pain and suffering, you say that you can't run it."

FEAC was asked to review the fusion programme in December, when the energy department was considering fusion budgets as low as \$200 million, and to present proposals for a range of programmes costing \$200 million, \$225 million, \$250 million and \$275 million respectively. Since then, however, the White House is said to have told Hazel O'Leary, the energy secretary, to put more money into fusion, and as a consequence, next month's budget proposal is expected to include a request for at least \$250 million.

As a result, the advisory committee declined to offer programmes at the lower two levels, saying merely that these would wreck the programme by endangering participation in ITER and forcing the closure of either DIII-D or C-Mod. It did offer a programme at \$250 million, and promised a much stronger programme if \$275 million were to be made available.

David Baldwin, senior vice president of DIII-D, endorsed the \$250-million proposal,

but told the panel that it was promising too much for \$275 million. "Your justification for the \$275 million is very weak," he said, adding that it was "incredible" to argue, as the committee had done, that an extra \$25 million would buy prolonged operation of TFTR, greater participation in ITER, and a stronger university science programme.

The FEAC report strongly endorsed continued US support for the design phase of ITER, while accepting that the United States is unlikely to be able to contribute much towards its construction.

Two members of the panel — James Thompson, a former director of PPPL, and Joseph Gavin, former president of Grumman Aerospace — dissented, arguing for a speedy withdrawal from ITER on the grounds that the country has no reputation left to lose as an international scientific partner, and that the machine itself will never be built. But the panel, led by Robert Conn of the University of California at San Diego, ruled that the ITER design phase should remain a top priority of the US programme — whether or not the machine itself is ever built.

The panel also called for another "review" of the US inertial confinement fusion efforts, the main alternative to magnetic fusion. But William Barletta, director of fusion at the Lawrence Berkeley Laboratory in California, where most of such research is done, accused Conn of "looking for a reason to assassinate" his programme. He expressed confidence that the Department of Energy would continue its support of the alternative work.

At the same time, Anne Davies, head of the Office of Fusion Energy at the DoE, said that the advisory committee was being optimistic about what could be achieved with \$250 million. She estimated that what it proposed at that level, including some operation of TFTR in 1997, would actually cost \$265 million.

"I don't think we can run TFTR next year if our budget is \$250 million," she said. At \$225 million, she added, "I don't think the programme will survive, because it will be so divisive. It will unravel." However, Krebs said that the panel had proposed a programme, "particularly at the \$250-million level", which she would support.

The structure of the US fusion programme — with three large domestic facilities chasing a shrinking pot of money while watching ITER drift to the top of the priority list — has led to bitter and occasionally public infighting in recent years, further undermining support for the field. "Something I can't over-emphasize is the need for community consensus," said Knotek. "The big message we received from Capitol Hill was that consensus is worth \$50 million to you," he added.

Colin Macilwain

Tokyo professor accused of perjury over evidence of HIV infection

Tokyo. A group of 72 people, including HIV-infected haemophiliacs, have filed an accusation of perjury with the Tokyo Public Prosecutor's Office against a scientist from Tokyo University who served in the Ministry of Health and Welfare in the early 1980s, and had held responsibilities for setting the ministry's policies on blood products and AIDS.

The group claims that Atsuki Gunji, a professor of public health, lied while giving evidence in a trial brought by the group against the government and pharmaceutical companies. Gunji testified twice in 1993, saying that ten years earlier he had been unaware that the infection routes for HIV are similar to those for hepatitis B.

A lawyer for the haemophiliacs says they have recently obtained a copy of a ministry document dated July 1983 that was submitted to the Ministry's AIDS study group stating that the infection routes of the two viruses were suspected to be similar. But Gunji reportedly maintains he never intended to lie, and refuses to accept that he did.

In October, the Tokyo and Osaka district courts urged the government and pharmaceutical companies to reach an out-of-court settlement with hundreds of HIV-infected haemophiliacs who have sued the government for the alleged late introduction of heat-treated blood coagulants and recommended a payment of 45 million yen (US\$420,000) to each plaintiff. The accusation filed by the group follows another one filed in April 1994 against Takeshi Abe, Gunji's former professor and head of the AIDS study group, who haemophiliacs have accused of "wilful negligence" resulting in death (see *Nature* 368, 680; 1994).

Troubled waters for ocean science

Washington. Republican divisions loomed large between the co-chairs of a joint hearing held last week on ocean sciences in the House of Representatives (see *Nature* 379, 283; 1996). Curt Weldon, chair of the research subcommittee of the National Security Committee, chastized Dana Rohrabacher, chair of the energy and environment subcommittee of the Science Committee, for creating "a false sense of savings" by cutting ships and programmes at the National Oceanographic and Atmospheric Agency (NOAA) which the Navy was left to pick up.

After the hearing, Walden said he would introduce legislation this year designed to improve co-ordination between NOAA, the Navy, and other agencies involved in ocean research.

Funding for US agencies extended

Washington. US President Bill Clinton signed yet another temporary spending measure last Friday (26 January) in a move that will allow funding for a number of agencies still without agreed budgets — including the National Aeronautics and Space Administration and the National Science Foundation — at last year's level until 15 March, almost the mid-point in the 1996 fiscal year.

Programmes which the House of Representatives wants to close, including the Advanced Technology Programme at the National Institute of Standards and Technology, will get 75 per cent of last year's funding level until then, in common with other programmes targeted by Republicans for major cuts (see page 386).

Megascience sets up task groups

Paris. The Megascience Forum, a body set up by the Organization for Economic Cooperation and Development (OECD) in 1992 to encourage intergovernmental collaboration in support for science, last week established three working groups. Two are on the specific areas of neutron sources and bioinformatics; the third is on general obstacles to cooperation, and includes subgroups on access to international facilities and legislative issues.

The forum has in the past produced reviews of big science projects, for example in oceanography, astronomy, deep-Earth drilling and neutron and synchrotron radiation sources. But last year, the OECD's high-level Committee for Scientific and Techno-

logy Policy extended the forum's mandate to allow it to set up working groups, made up of government officials and scientists, to propose new cooperative agreements.

"Instead of looking at broad issues, we will now be looking at more specific issues," says Stephan Michalowski, its executive secretary. "We will no longer only be talking, but proposing solutions", adds Peter Tindemans, the Dutch chairman of the forum, who says that the forum rejected proposals for a working group on seismology because it felt that the necessary scientific consensus on what is needed was lacking.

'No' to commercial fossil collecting

Los Angeles. An opinion poll of 300 US citizens has revealed widespread opposition to the commercial collecting of fossils found on public land. More than 90 per cent of respondents to the poll, which was conducted last year by MKTG Inc., a New York-based polling service, said they would report to "appropriate scientific authorities" any fossil found on public or private land.

In addition, over 80 per cent said they would allow a museum or university to collect such a fossil, once found. A similar proportion expressed the view that it should be illegal to collect fossils found on public land, sell them, destroy them or export them.

Democrats thank President Clinton

Washington. The Democratic R&D Task Force in the House of Representatives, set up last autumn by George Brown (Democrat, California), has written to President Bill Clinton to thank him for showing some backbone in recent negotiations over the science budget (see *Nature* 378, 229; 1995).

Brown, who expressed doubts last summer over whether Clinton would stand and fight for research programmes, says that he is now impressed with the president's refusal to accept Republican cuts to climate change research and the Advanced Technology Programme. Clinton has vetoed the appropriations bill containing these cuts.

French risk body suspends work

Paris. A commission set up by the French government to assess industrial and technological risks has suspended its activities in protest at what it describes as uncertainty over the government's willingness to maintain an 'independent institution' with power to produce opinions on its own initiative. Over the past year, the government has not replaced members of the Collège de la Prévention des Risques Technologiques (CPRT) whose terms of office have expired, and its apparent lack of interest has fuelled rumours that it intends to abolish the CPRT (see *Nature* 379, 4; 1996).

But Corinne Lepage, France's minister of environment, has now taken up the CPRT's cause and asked for it be reconstituted within her ministry. Although this move would downgrade the CPRT, which is currently attached to the prime minister's office, it has been cautiously welcomed by the group itself — partly because Lepage is an outspoken environmentalist. The CPRT will continue to suspend its activities, however, until it gets firm guarantees that its independence will be preserved, and that it will be given an adequate budget. Otherwise it will resign *en bloc*, says one member.

South African tribunal postponed

Cape Town. The tribunal which was due to begin hearings this week investigating allegations against William Makgoba, deputy vice-chancellor of the University of the Witwatersrand, has been postponed after the withdrawal last week of one of its members, Walter Kamba of the University of Namibia (see *Nature* 379, 199; 1996).

Robert Charlton, vice-chancellor of the university, said in a statement that in creating a new timetable for the adjudication of the allegations, the university council had agreed "that every effort would be made to constitute a new panel that meets legitimate concerns that Professor Makgoba may raise". The two remaining members, Lord Flowers, a former vice-chancellor of the University of London and Sir Colin Campbell, vice-chancellor of the University of Nottingham, would still serve on the tribunal, subject to their availability in the light of other commitments.

Gene tests: who benefits from risk?

The insurance industry is likely to find it more difficult to use genetic data on disease susceptibility than some of its critics claim. But it is already contesting demands for strict controls on the way this data is used.

GENETIC testing is one of the fastest-growing areas of medical diagnostics. In recent years, tests have been devised for Huntington's disease and cystic fibrosis. A test was launched for Alzheimer's disease last week and an announcement is expected later this year on a test for the breast cancer gene *BRCA1*. The now-defunct US Office of Technology Assessment estimated a tenfold rise in the numbers of genetic tests over the next decade.

But genetic testing continues to raise social, legal and ethical issues. A key question is whether genetic information should be made available to health insurance and life insurance companies — and whether they should be allowed to ask prospective policy-holders to undergo genetic tests as a condition for taking out an insurance policy.

Opinions are divided. Certain ethics groups, such as the US National Institutes of Health-Department of Energy Working Group on Ethical, Legal, and Social Implications (ELSI) of the Human Genome Project, some politicians, and sections of the public maintain that genetic information should not be given to insurance companies.

Others, such as the Genetics Interest Group (GIG) in London, an umbrella organization of more than 100 charities representing people affected by genetic disorders, agree with life insurers which say that such a stance is unrealistic, especially if genetic testing becomes widespread. But, whereas the GIG would like the use of gene tests by life insurers to be regulated by national legislation, some insurance companies would prefer to do the regulating themselves.

Evolving legislation

The current situation regarding legislation remains mixed. In the United Kingdom, Germany, Japan, Italy, Spain and Portugal, there are no laws governing the use by insurance companies of genetic tests or their results. In the United Kingdom, companies represented by the Association of British Insurers (ABI) do not at present make genetic tests a condition of issuing policies, but reserve the right to have access to the results of previous tests, according to Paul Smee, an ABI spokesman.

A similar situation exists in Germany, with the exception that only health insurance companies — not life offices — ask to see the results of previous genetic tests for policies that exceed DM250,000 (US\$174,000). Siegfried Ackermann, chief medical officer for Allianz insurance, says

genetic testing is still very expensive and insurance staff can calculate premiums accurately using existing medical information, such as questions about family history.

Other countries, such as France and the Netherlands have imposed moratoria — temporary prohibitions — before a final decision is made. In France, genetic information can be used only for medical and

It could be you: the use of DNA sequencing data is becoming a key issue in insurance.

research purposes, but regulations do not specifically address the question of genetic testing by insurance companies. The French Federation of Insurance Companies, however, has agreed to a five-year moratorium on the use of genetic information.

In the Netherlands, however, life insurers can require a genetic test for policies above NLG300,000 (US\$224,000). But a bill now going through parliament aims to end this practice by restricting insurance companies to being allowed access to the results of previous genetic tests but not to ask for a test. Belgium, Austria and Norway, meanwhile, have opted for an indefinite ban on both.

In the United States, ten states prohibit genetic testing and the use of genetic information for health insurance purposes. Private health insurance is widespread in the United States as only 11.7 per cent of the population qualifies for public health insurance, according to Richard Coorsch, a spokesman for the Health Insurance Association of America. He says its 220 member organizations do not require genetic testing, nor do they ask to see test results.

There are no restrictions on US life insurers requesting genetic tests. Life insurers have successfully lobbied to ensure that companies are not prohibited by state laws from having access to genetic information. But Kenneth Vest, a spokesman for the

American Council of Life Insurance (ACLI), says his organization's members do not ask prospective policy-holders to take tests. The ACLI withdrew from a task force on genetics and insurance set up in 1993 by ELSI when the final report urged insurers to consider a moratorium on the use of genetic information. The Health Insurance Association, by contrast, took a neutral stance.

There are two main reasons why life offices want access to the results of genetic tests. Life insurance works on the basis of estimating how long a person will live (see page 391). If a genetic test can improve this estimate, it will provide more accurate calculations of insurance premiums, says Vest.

Adverse selection

André Chuffart, a vice president of the Swiss Reinsurance company in Zurich, argues that until genetic testing became an option, someone whose parent suffered from Huntington's disease would have difficulty obtaining life insurance. Using genetic testing and/or information, says Chuffart, will enable many who do not show a predisposition to obtain cover at standard rates.

The second reason why insurers want access to the results of genetic tests is as a defence against what insurers call 'adverse selection or anti-selection' — the fact that, all other things being equal, people who know they are at high risk will be more likely to seek insurance. Insurance works on the principle that a policy-holder pays according to the degree of risk he or she brings to the insurance fund. So high-risk individuals pay more, low risks pay less. "Where applicants know their future risks [for example, from a genetic test] and insurers do not, this can lead to the breakdown of the entire insurance market," says Paul Brett, an actuary based in Cologne.

Adverse selection is one reason that US health insurance companies have not lobbied as strongly as life insurers for access to genetic data, according to Coorsch. Whereas life insurance policy-holders tend to fund premiums themselves, 90 per cent of health insurance policies in the United States are ►

Sinclair Stammers/SPL

On other pages

■ Limited powers of prediction	390
■ What the public thinks	391
■ Speeding up the process	392

Complexity limits the powers of prediction

GENETIC tests can already be used to predict with relative ease the incidence of disorders that are associated with a single identified mutation in a single gene. But few tests exist for disorders caused by mutations in several genes, or for diseases brought on by interactions between an individual's genes and his or her lifestyle or environment — precisely those which, because they cover the most common diseases, are also likely to be of most interest to insurance companies.

A typical example of a disease whose outcome can be predicted relatively easily is Huntington's disease (HD), a fatal and incurable neurodegenerative disease that affects around 1 in 20,000 of the European population during mid-life. HD is associated with a single identified mutation in a single gene. Sufferers have many repeats of a trinucleotide sequence (cytosine-adenine-guanine) within a particular gene. Researchers have identified both the gene involved and its chromosomal location, and can predict with accuracy whether a carrier will develop the disease.

But such diseases tend to be the exception. The nature and location of the genes responsible for other disorders are often far from clear. Identifying the nature and implications of mutations of the large gene whose malfunctioning is associated with cystic fibrosis (CF) is more complicated.

CF affects approximately one in 2,500 of the Caucasian populations of North America and Northern Europe. Seventy per cent of sufferers carry a particular, known mutation in the gene that encodes the protein responsible for the disorder. But a negative test for this mutation does not guarantee freedom from disease. This is because a potential CF sufferer could be carrying any one of the more than 500 other known mutations within the same gene that are also associated with the disease.

Furthermore, cystic fibrosis is still relatively well-defined. Predicting more common disorders such as breast cancer and cardiovascular disease, which are more likely to be of interest health and life insurance companies, will be more challenging, as they involve several different genes.

For example, mutations in five separate genes are known to increase an individual's susceptibility to colon cancer. Similarly, mutations in at least two genes are known to predispose to breast cancer. But the absence of such mutations does not necessarily mean that the person who has been screened will not develop either disease.

Take breast cancer. One in 10 women living in Western societies is likely to develop breast cancer by the age of 85 and 25 per cent of these sufferers will die from

the disease. About 5 per cent of breast cancer cases are hereditary. But mutations in the two known susceptibility genes — *BRCA1* and *BRCA2* — account for only 80 per cent of all the hereditary cases.

As a result, genetic test for *BRCA1* and *BRCA2* will not be of any value to the remaining 20 per cent of women predisposed to breast cancer, and will be even less informative about the 95 per cent of cases that are not hereditary.

The designers of genetic screening tests

has yet to be refined.

The genes that predispose to late-onset diabetes, for example, which is characterized by resistance to insulin, have not yet been identified. But Leif Groop, a geneticist at the University of Lund in Sweden, who is conducting a study in families in Norway, predicts that between two and ten such genes may exist.

Even when they are identified, there will still remain the daunting task of interpreting the way these genes predispose to a disease that is strongly influenced by external factors such as poor diet and obesity.

Progress in identifying the genes that may influence cardiovascular disease remains equally disappointing. Even if such genes are eventually identified, it will still be difficult to translate this information into the quantitative prediction of risks.

In the United States, a quarter of men and women under the age of 65 and about half of those over 65 die from diseases of the cardiovascular system, according to the National Center of Health Statistics. In some cases, such diseases are the inevitable results of a particular mutation; one such mutation, for example, causes long QT syndrome, a disorder characterized by sudden cardiac failure. But most other forms of heart disease are due to a combination of genetic susceptibility and lifestyle.

Even in familial hypercholesterolaemia, which — like long QT syndrome — is caused by a defined mutation (in this case in the cholesterol-carrying LDL protein) and invariably leads to disease, the onset and severity of the condition can be affected by diet. And the majority of genetic effects are even more difficult to use for making predictions about an individual's health.

In atherosclerosis, for example, a major cause of heart attacks and strokes in the developed world, some patients are known to have mutations in the genes that encode proteins required for fatty substances to be processed normally in the blood. But atherosclerosis is a complex disease involving a wide range of molecules, while many non-genetic factors, including diet, alcohol, stress and smoking, modulate the risk of developing this disease.

Even a sophisticated knowledge of genetics may not allow doctors to predict with any degree of certainty the likelihood of contracting polygenic and multifactorial diseases. Nevertheless, this knowledge will include a mass of data that may influence an insurance company's perception of an individual's risk. □

Alison Abbott, with additional reporting by Barbara Cohen and Rory Howlett

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Testing times: the more genes involved, the more difficult prediction becomes.

are faced with an additional difficulty, the fact that breast cancer researchers have not yet unravelled the mechanism by which the protein produced by the *BRCA1* gene appears to prevent cancer from developing in normal cells.

A genetic test — expected to be marketed shortly by Myriad Genetics in Salt Lake City, Utah, — will involve the determination of a DNA sequence for *BRCA1*. The gene is very large; at 7.8 kilobases it is about 10 times the length of a normal gene.

Screening tests may reveal a number of sequence alterations. Some will effect the production of proteins whose presence or absence may be responsible for the disorder. The effects of others, however, remain unknown, making it impossible to use information about the mutation to predict without a functional test whether or to what extent someone is at increased risk.

Genetic information about disorders of this type presents the greatest challenge to researchers as, in addition to the existence of predisposing genes, an understanding of the powerful influence of environmental factors on the development of such diseases

BSIP Laurent/SPL

provided by employers who pay around 70 per cent of the premiums. Moreover, the health of individuals is less relevant to employers' groups responsible for managing company health policies as premiums are not set by assessing each policy individually, but by a long-term 'community' average in which low risks subsidize the higher risks.

Nonetheless, Coorsch believes health insurers will agree to relinquish access to genetic information only if they are allowed to continue using other medical information, such as a family history of disease, especially for policies that bring higher risks.

Life insurers agree, but add that genetic tests should be treated in the same way as other medical information, says Spencer Leigh, an actuary and chief underwriter for Royal Insurance in Liverpool, the United Kingdom, who presented a paper, "Freedom to underwrite", at a meeting of the Institute of Actuaries in London last month.

A portfolio of problems

Critics such as Peter Harper, chairman of the Royal College of Physicians' Clinical Genetics Committee, cite various reasons why they disagree. First, they argue that information about predisposition to a disorder is not the same as medical history because such information, unlike certain genetic tests, does not indicate how long a person will live.

Second, they argue that people should not be denied insurance — or charged higher premiums — for a condition for which they are not responsible. Third, there are doubts whether the insurance industry will be capable of handling and interpreting sensitive and complex information.

Fourth, the prospect of increased insurance premiums — or discrimination in employment — may deter people from undergoing genetic tests, thus affecting the impact of genetic testing programmes. Fifth, use by insurance companies of genetic data, for example to up-rate a policy, will violate the wish of someone who has chosen not to know the result of a genetic test.

Other observers, such as Hugh Watkins, newly appointed professor of cardiovascular medicine at the University of Oxford, question the value to underwriters of polygenic data — particularly for disorders such as cardiovascular disease — as the research in this area is "in its very early stages".

Groups such as the UK Nuffield Council for Bioethics, which published a report on the ethics of genetic screening in 1993, add that they recognize the insurance industry's concerns and thus advocate a moratorium — rather than an outright ban — on insurance companies using genetic information.

Many insurance industry officials oppose a moratorium. Some, such as Leigh, regard it as effectively a ban. But they also dispute the reasons advanced for denying them access to genetic information.

Chuffart, for example, who is the author

of a paper entitled "Genetic underwriting", says premiums might not necessarily rise significantly for higher risks if insurers were allowed to use genetic information. "Competition among offices is so intense that there will always be some company willing to offer lower rates," he says.

The charge that it is unfair to put higher premiums on high genetic risks represents a fundamental misunderstanding of the principles of insurance, says Leigh. Chuffart agrees and says companies are businesses and not "social welfare organizations". Chuffart concedes, however, that policyholders would need to be counselled if they were found — against their wishes — to be predisposed to a disease.

Leigh suggests personal circumstances should not be allowed to interfere in business. In motor insurance "someone who lives in a high-risk area is going to pay more" even if he has no option but to live there because he cannot afford to move. "Should he pay less for car insurance because the address isn't one of his choice?" he asks. "It is not the insurance companies that are being unfair, it is life itself."

One of the obstacles here is the absence of quantitative data to substantiate the views of the protagonists. There is very little academic work on genetic testing and life insurance. The insurance industry itself has only recently begun to address it in detail.

On the scant evidence available, it is becoming apparent that insurance companies could be persuaded to ignore polygenic data on the grounds of cost. At the moment, the industry has a well-honed system of estimating life expectancy and mortality, based on tables that are continually updated

by actuarial associations around the world. These tables would have to be recalculated at significant cost if polygenic information were to become the basis for premiums.

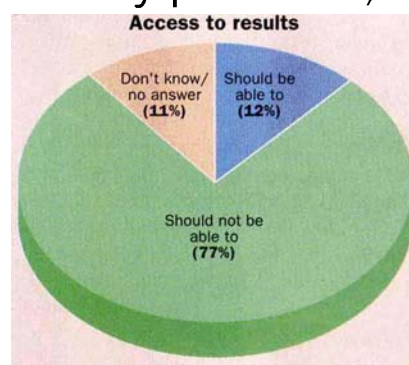
The method of assessment would also have to change. Life insurers do not assess each policy individually, as in motor insurance, but have a standard premium proportional to the probability of death and amount of insurance needed, plus administrative expenses, added to a profit. Companies usually 'load' a premium by a small amount to cover the unpredictable. As a result, 95 per cent of applicants are offered insurance at standard rates, with few medical questions and little administrative work for the office.

Economic implications

Sufferers from monogenic disorders, such as Huntington's disease, fall within the 5 per cent of cases that have to pay more or are denied policies. But sufferers from polygenic disorders, such as heart disease, are likely to be spread within the majority 95 per cent. So if insurance companies want to use polygenic data, they will have to assess each policy individually. "Genetic information will become uneconomic if each applicant has to have an individually tailored policy," acknowledges Paul Smee, head of life insurance at the ABI.

Insurers, however, insist that the potential losses from adverse selection override the fact that genetic information may be of little actuarial value. If data remain off-limits, companies may have to raise premiums to compensate for adverse selection, says Robert Pokorski, vice president of medical research for Swiss Re America, and a ►

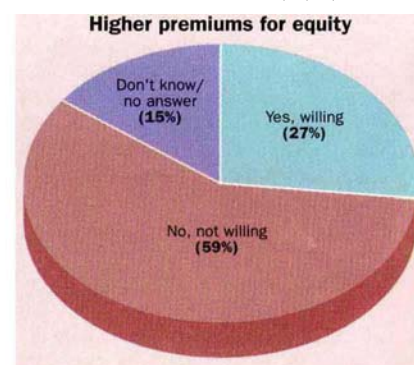
Privacy preferred, but what will it cost?



A clear majority of people do not want insurance companies to have access to the results of genetic tests, according to US public opinion surveys conducted for the American Council of Life Insurance (ACLI).

Most respondents seem to know little about genetic testing, but 77% of those surveyed in 1994 said that life insurance companies should not be allowed access to results (above left).

Individual policy-holders proved



reluctant to pay more for their policies in order to ensure that life cover would be available to everyone at the same rate, regardless of their health and the risk they represent (above right). Only 27% said they would pay more to make a universal rate possible; more than twice that number would not.

The surveys also indicate that the insurance industry "continues to be held in low esteem by the public"; 29% viewed it favourably in 1995. □

DNA chips intensify the sequence search

MANY of the current difficulties in interpreting genetic data are due to technology as much as biology. The first problem facing those hunting down mutant genes is knowing just where to start looking. But the second — and in some ways more formidable — problem is that current analytical techniques cannot sift data fast enough to predict complex, late-onset diseases.

Predisposition to most diseases is affected not by a single gene, but by several. Conversely, each disease-related gene can have many possible mutations. Each mutation therefore has to be searched for individually. The sheer numbers of analyses required are beyond the scope of tests based on sequencing or hybridization technologies that are now in use.

But this whole situation could be changed by so-called 'DNA chip' technology, possibly within a decade. This technology essentially allows large numbers of genetic mutations to be identified simultaneously. It works by synthesizing short

segments of nucleic acids that correspond to the gene under investigation and binding them directly on to small squares of wafer-thin glass chips.

The chips with their attached segments — known as oligonucleotide probes — are then incubated, together with the copies of the patient's gene being tested. If the patient's gene has the same sequence as the probe, it will bind exactly and be detected. Mismatching sequences, indicating mutations, can be identified automatically by advanced software.

The power of DNA chip technology over conventional sequencing techniques lies in the large number of probes that can be squeezed onto a single chip. The company leading this field, Affymetrix, based in Santa Clara, California, is already using chips that can fit 16,000 different oligonucleotides on to a chip with an area of less than 1.5 cm², and has developed a technique for increasing this by a factor of 25.

Affymetrix, 65 per cent of which is now

owned by Glaxo, (see *Nature* 373, 372; 1995), has already developed chips that can, for example, check for drug-resistance viruses in patients with HIV. It is also developing a chip that can screen the cystic fibrosis gene for 80 per cent of all known mutations simultaneously.

The company is now planning to address the more challenging problem of multigene disorders. This is already exciting the interest of some life insurance companies, as these types of disorders are far more common among those applying for health and life insurance. To streamline the very complex process, Affymetrix plans to perform multigene screening by examining profiles of gene expression, rather than directly looking for gene mutations, using chips containing oligonucleotides complementary to messenger RNA rather than DNA.

The company claims that such a system for screening for colon cancer, for example, could be ready for testing within five years.

Alison Abbott

member of the ACLI Genetic Issues Committee.

But evidence that people would rush to buy large amounts of insurance, given exclusive knowledge of risk remains unclear. Some data compiled by the Continuous Mortality Investigation Bureau in the United Kingdom — used as the basis for UK mortality tables — shows that the death rate rose by up to 66 per cent in a group that took advantage of a life insurance scheme offered during the mid-1980s in which a medical questionnaire was not required.

Indeed, the practice of many life insurance companies to offer policies — even for those in middle age — requiring no medical information suggests this does not lead to losses. The reason, according to one financial adviser, is that companies know that a high proportion of the over-40 age group is already insured. "They are trying to attract the few who have no insurance. They know there will not be a flood" of applications.

Faced with the continued prospect of regulation — especially after the expiry of moratoria — insurers are starting to take the debate forward, themselves. In the US, ACLI will meet on 10–12 February for a two-day seminar in Atlanta on genetic testing. The meeting will focus on developments in the science of genetics, actuarial implications and political developments.

In Europe, a similar discussion will take place in September in London organized jointly by the UK Institute of Actuaries and the Royal Society. The ABI, meanwhile, is independently preparing an industry position paper on genetic testing, although no publication date has been fixed. The British government decided not to impose a dead-

line for submission — a House of Commons committee had recommended one year from July 1995 (see *Nature* 379, 195; 1995).

So far, two approaches have been proposed. The first, put forward by Nicholas Barr, a consultant to the World Bank and lecturer in the economics of insurance at the London School of Economics, assumes a moratorium on the use of genetic data by insurers. Policies would be issued at standard rates, but if a policy-holder succumbed to a 'genetic' disease — according to a list kept by an appropriate authority — the sum of money paid out should not exceed a ceiling specified at the time of the contract and would come from an industry-wide levy imposed on premiums for listed disorders.

Policy proposals

This option, supported by the UK House of Commons Select Committee on Science and Technology, does not appear popular with the insurance industry. The ABI has said it is unworkable. Leigh described it as "ludicrous" because the ceiling would penalize high-value policy-holders who happen to die from a disease on the 'list'.

It might also lead doctors to leave open the cause of death if the deceased's original sum assured is more than the official entitlement, he adds. A list restricted to genetic diseases would also be controversial. "How would you decide which disease to include?" asks Bowley. "If I was a severe diabetic, I would want to be included." Paul Smeed, in addition, says the industry would prefer the taxpayer to fund any levy, although he does not rule out it having to pay. Industry-wide pooling is not favoured by companies as it limits the scope for competition.

Leigh has an alternative of his own. Insurers should be allowed to see test results and take them into account for large policies, he says. But these results "will be ignored for the purposes of life insurance where the initial sum assured on death is £50,000 or below; the term 10 years or less." However, he adds that the term 'genetic' will need to be defined so as not to include cholesterol, ECG, blood or X-ray tests, so that insurance companies can continue underwriting policies as at present.

Organizations in the United States have also suggested that companies could offer specially tailored products, such as term-limited insurance at favourable prices, in return for proof of health maintenance.

The UK industry's response to Leigh's suggestion has not been encouraging. The ABI appears lukewarm about the idea of an artificial ceiling. And at least one industry representative has said that £50,000 may be too large a sum for the smaller companies.

Genetic interest organizations both in Britain and the United States, however, say they will not oppose such policies, providing the industry can demonstrate it is equipped to interpret the data accurately and use it responsibly. Most large offices already call on the services of eminent consultants as chief medical officers. A minority also consult specialist geneticists. But smaller outfits have no medical provision at all. Chuffart says a code of conduct will need to be drawn up to guard against abuse from both sides. □

Ehsan Masood, with additional reports from Sally Lehrman in San Francisco, Quirin Schiermeier in Munich, Declan Butler in Paris and Richard Nathan in Tokyo.

RGS, Shell and Nigeria

SIR — The headline to your News story about relations between the Royal Geographical Society (with the Institute of British Geographers) and Shell — “UK geographers vote to cut links with Shell (*Nature* 379, 104; 1996) — gives a wrong impression. Some geographers voted thus, but by no means all.

We have more than 13,000 fellows and members. Of these, 900 were at our academic annual conference in Glasgow, of whom fewer than 200 attended the open meeting that discussed Shell's corporate patronage of the society. In fact, 156 fellows supported the motion to sever links with Shell. That is only 1.1 per cent of our fellowship.

I will not comment on Shell's environmental and political record in Nigeria until after we have learned more from the forum planned here later this year. I would imagine, however, that Shell would welcome a return to democratic rule in that unfortunate country as much as we all would.

I can add to what you say about Shell's record as a benefactor of research. The company has been a generous supporter of our work for many years. It sponsors a department that gives training and advice to people going on expeditions; it subsidized a prize-winning schools teaching pack on rainforests; it paid for a purpose-built rainforest research station in Borneo; and it has often helped our scientists in the field.

This patronage has been entirely disinterested. Shell has asked for nothing in return. It was our idea to call the company Corporate Patrons, alongside three other British companies. Shell scarcely mentions this support: it appears only on page 14 of its leaflet *Shell in Society* among hundreds of other good causes the company helps.

John Hemming

Royal Geographical Society,
1 Kensington Gore,
London SW7 2AR, UK

Not so black

SIR — In the leading article “Europe's black sheep”, you write: “Greece (where Aristotle used to live) is at the bottom of all lists; total spending on research and development runs at about 0.5 per cent of gross domestic product” (*Nature* 378, 222; 1995). In this context, the parenthetic phrase is polemic, ironic and — to use your own words — “to some degree unfair”. Furthermore, this phrase does not place Greece deeper in “all lists”. So there is no additional message. More importantly,

however, this statement creates doubts about your journalistic objectivity. Scientists who live and work in Greece and who publish their research results in international peer-reviewed journals (including *Nature*) will not be amused.

Research done in Greece compares well with that in Italy and other larger member countries of the European Union (EU). And in order to put Professor Antonio Ruberti's proposal in its proper context, it would have also helped if, as well as publishing the percentages of gross domestic product spent on research and development in EU member countries, you had shown other related indices. For instance, Greece has the highest military spending per capita of all NATO countries, mainly because of its strategic position and the requirements of the recently ended Cold War. Greece has had to buy military equipment from Britain and France, among others, and support in this way the military research in those countries, on which as you indicate, these two countries “spend substantial sums...”.

Athanassios Giannis

Institute for Organic
Chemistry and Biochemistry,
University of Bonn,
Gerhard-Domagk-Strasse 1,
D-53121 Bonn, Germany

Clinical trials

SIR — Recent correspondence on xenotransplantation (*Nature* 378, 434; 1995) has demonstrated the need, in the words of David White, for “plans for a clinical programme dealing with such issues as disease control, ethical considerations and regulatory affairs [to] be initiated as soon as possible”.

Such plans have in fact been initiated. In the United Kingdom, the Nuffield Council on Bioethics will publish on 6 March a report on xenotransplantation that has been in preparation since early 1995. The Department of Health has established a committee under Professor Ian Kennedy to address the issues; a report is expected in mid-1996. In the United States, the Institute of Medicine is producing a report on xenotransplantation and guidelines on disease control are shortly due from the Public Health Service.

It is to be hoped that “the final decision on when to commence clinical trials will be based” not only on “the experimental data” but also on public discussion and on the development of a regulatory framework that is rapidly being put in place.

David Shapiro

Nuffield Council on Bioethics,
Nuffield Foundation,
28 Bedford Square,
London WC1B 3EG, UK

No way back?

SIR — We endorse Pere Puigdomènech's opinion (*Nature* 378, 126; 1995) that Spanish science is on the brink of a crisis. Until a few years ago, postdocs returning from a period abroad could easily find a position within the research establishment, either at a university or at the Spanish Council for Scientific Research (CSIC). That seems no longer to be the case, at least for most young scientists. The situation is especially worrying in the universities, which account for almost 80 per cent of Spanish research. The strenuous efforts in many Spanish universities in the past few years to improve the quality of science and the training of new scientists are now in jeopardy.

Although the Socialists increased research funding during their early years in government, they were at the same time sowing the seeds of the present problems. The Socialist government started to clog up the university system early in 1984 when it convened a one-off extraordinary board of evaluation (called *Idoneidad*, Spanish for ‘suitability’) to make appointments to faculty positions. In many cases, people who could barely fulfil the legal requirements (little more than a PhD and some teaching experience) became university professors, regardless of their scientific ability.

And during the budgetary expansion of the mid-1980s, positions were also being filled by the regular Spanish system of evaluation known as *Oposiciones*, which does take into account the scientific productivity of the candidate. And in Spain, a professorship is a permanent position with the status of a civil servant appointed for life.

Many university departments are now staffed by people with an average age of 40 to 45. With the compulsory retirement age for academic staff recently raised to 70, few positions will become available in the next 25 or 30 years. Moreover, there are no non-tenured professors or researchers working under contract in universities. For postdocs on a three-year contract paid by the government after their return from abroad, there is no opportunity for the renewal of their contracts anywhere in Spain. So very few of the postdocs now returning will ever have access to a faculty position within a university department. To make things even worse, universities are producing an increasing number of PhDs just to keep up their funding levels and scientific productivity.

As two postdocs with their three-year appointments about to expire we wonder if emigration is once again the only future for Spanish scientists.

Yolanda S. Lopez-Boado
Fernando Segade

Departamento de Bioquímica
y Biología Molecular,
Universidad de Oviedo,
33006 Oviedo, Spain

Birth dates in Florence...

SIR — Recent correspondence showing a constant asymmetry in the birth-date distribution of students entering the Porto faculty of medicine¹ has led us to investigate whether a similar asymmetry would emerge in the birth-date distribution of the students entering our faculty of medicine in Florence. Similar premises apply to Italian students as to Portuguese.

our faculty.

We considered the birth dates of the 957 students who had entered the Florence faculty of medicine in the five academic years from 1988 to 1992 and who were still registered on 1 January 1994. The analysis revealed the same asymmetry, constant during the five-year period.

FIRENZE FACULTY OF MEDICINE — STUDENTS IN BIRTH-DATE QUARTERS				
Year of entry	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
1988-89	40 (46.7)	63 (47.0)	46 (48.1)	36 (44.2)
1989-90	37 (45.4)	59 (46.4)	35 (46.7)	51 (43.4)
1990-91	42 (48.9)	61 (49.5)	49 (51.2)	44 (46.4)
1991-92	49 (49.4)	58 (50.3)	47 (52.3)	45 (46.9)
1992-93	52 (48.0)	55 (49.5)	45 (50.8)	42 (45.7)
Total	221 (238.6)	296 (242.7)	222 (249.1)	218 (226.6)

Starting from the 1988-89 academic year, the *numerus clausus* (programmed number) was introduced in almost all Italian faculties of medicine. The main result has been that entrants to the medical faculty are recruited from among the most successful high school students, due to the scarcity of places as compared to the number of interested applicants. Last year, for instance, 547 applicants competed for the 200 places available in

Our data show a significantly higher incidence of students born during the second trimester of the year. Although our figure is not so important as that found in Portugal, it is still significant ($\chi^2 = 16.28$, $P < 0.005$).

The dates of birth of our students, subdivided into trimesters, are shown in the table. The expected number of students is given in parentheses, taking into account the Italian birth data for the

same five years². These data too present certain asymmetries, though not so significant as the others. It may be noted, for example, that in the 12 years for which the Italian birth data have been taken into account (1965-78) December was always (with two exceptions, which were November) the month of the year with the lowest birth numbers.

We concur with Azevedo *et al.* that the hypothesis of Michael Holmes³, on the importance of the seasons in the early development of the child, is very attractive, and we believe that the fact that "the self-wiring of the brain" is dependent on external stimuli⁴ might give weight to this hypothesis.

We hope that someone in the Southern Hemisphere will now analyse similar data, in order to confirm (or to confute) the 'spring advantage'.

Vieri Boddi

Enzo Brizzi

Antonio Conti

Gian Franco Gensini

*Faculty of Medicine,
University of Firenze,
I-50134 Firenze, Italy*

1. Azevedo, I., Pinto -do-Ó, P. & Borges N. *Nature* **376**, 381 (1995).

2. *Nascite e decessi — Anno 1991*: ISTAT (Istituto Nazionale di Statistica) Roma, 1994; *Statistiche demografiche — Anni 1986, 1987, Tomo 2; Parte prima, Nascite e decessi*. ISTAT, Roma, 1988.

3. Holmes M. *Nature* **373**, 468 (1995).

4. Stryker M.P. *Nature* **375**, 277-278 (1995).



...meanwhile, elsewhere in Italy

SIR — Success in sports, high-school performance and scientific creativity seem to correlate with birth dates¹⁻⁴. Recent observations by our group suggest that genetic and developmental parameters in newborn infants bear a relationship with season of birth.

The search for a possible adaptive

the population of Penne, a small rural town 438 metres above sea level in south-east Italy. The majority of the people admitted to the Town Hospital (including puerperae) live between 400 and 500 metres above sea level, but there are villages located at higher altitude. Cold winters characterize the climate of this area.

ACP₁*A GENE IN A SAMPLE OF NEWBORN FROM THE POPULATION OF PENNE

Month of birth	ACP ₁ *A (%)	Gestation duration (weeks)	Birth weight (g)	Sex proportion (%)
Oct./Nov.	35.9±3.7	39.4±0.13	3,272±56	51.8±5.4
Dec./Jan.	35.5±3.5	39.5±0.15	3,345±51	51.6±5.2
Feb./Mar.	25.0±3.2	39.7±0.16	3,429±46	45.7±5.2
Apr./May	31.6±4.7	40.1±0.07	3,423±64	47.9±7.2
Jun./Sept.	19.7±4.9	39.6±0.21	3,383±61	45.6±8.7

value of ACP₁ (acid phosphatase controlled by locus 1) was the main purpose of our survey. The enzyme is a phosphotyrosine phosphatase and may have important functions in cellular growth regulation and in the modulation of glycolytic rate⁵. Population studies suggest a role in adaptation to environmental temperature⁶ and we have recently shown associations with obesity and diabetes^{7,8}.

We studied 352 newborn infants from

The table shows ACP₁*A frequency, gestational length, birth weight and sex of newborn infants from the population of Penne. ACP₁*A, the allele associated with the lowest ACP₁ activity, gestational length and birth weight show a significant association with season of birth ($P<0.01\%$). Most seasonal heterogeneity is due to differences between infants born in October–January and infants born in February–September, corresponding to conception in January–April

and May–December respectively. Infants born from October to January show a higher frequency of ACP₁*A, a reduced gestational duration and a lower birth weight compared with infants born from February to September. This pattern is in line with population and biochemical studies suggesting that low-activity genotypes may have an increased metabolic rate and a relative advantage in cold environments⁶.

Our observations suggest that season of conception and/or environmental temperature may influence intrauterine development. The effects on neonatal parameters measurable at a general population level seem rather small, but may be far-reaching in selected groups of talented people.

F. Gloria-Bottini

N. Lucarini

M. La Torre

*Department of Preventive Pediatrics,
University of Rome Tor Vergata,
School of Medicine,
00173 Rome, Italy*

1. Dudink, A. *Nature* **368**, 592 (1994).
2. Edwards, S. *Nature* **370**, 185 (1994).
3. Holmes, M. *Nature* **373**, 468 (1995).
4. Azevedo, I., Pinto-do-Ó, P. & Borges, N. *Nature* **376**, 381 (1995).
5. Stefani, M. et al. *FEBS Lett.* **326**, 131–134 (1993).
6. Ananthakrishnan, R. & Walter, H. *Humangenetik* **15**, 177–181 (1972).
7. Paggi, A. et al. *Dis. Markers* **9**, 1–7 (1991).
8. Gloria-Bottini, F. et al. *Experientia* (in the press).

Automatically determine HLA types with tools from an ÅÄÖ country?

Can you say your method of determining HLA types is completely accurate? If not, a new approach to detecting HLA types offers you real accuracy.

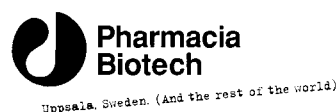
Introducing HLA SBTyper™, the new software for automatic sequence based typing from Pharmacia Biotech in Sweden—an ÅÄÖ country.

HLA SBTyper uses raw sequence data, generated by ALFexpress™, to automatically generate high-resolution HLA typing for class II genes. It determines HLA types by comparing sequence results against a database of well defined and established subtypes. And it all takes just two minutes. ALFexpress

and HLA SBTyper also let you access raw data—so you can always go back and verify your results.

HLA SBTyper with AutoLoad™ Solid Phase Sequencing Kit and ALFexpress—together they're the tools that present you with the most accurate way of determining HLA types, and handle every step from capture of PCR products to evaluation.

Just call us at 1 (800) 526 3593 in the United States or +46 18 16 5011 from the rest of the world. Ask for more information about what we Swedes have done for automated sequence based HLA typing. You'll find we've taken every å (pronounced ooh) ä (aah) and ö (uh) out of determining HLA types.



Inexpressible trinucleotides

SIR — It is indeed interesting, as Charles Jennings says in his News and Views article¹, that the mutant polypeptide encoded by expanded-repeat genes in Huntington's patients is associated with a specific protein, and the immunological evidence for a novel structure in the expanded-repeat proteins is highly suggestive.

But, while these observations may well be relevant to the pathology in Huntington's disease, in our view they cannot provide a general explanation for the trinucleotide repeat diseases. This is because the commonest trinucleotide diseases are fragile X syndrome, which is caused by expansions of CGG and CCG trinucleotides, and myotonic dystrophy, which is caused by expansion of a (CTG)_n tract. Neither in fragile X nor in myotonic dystrophy is the trinucleotide repeat expressed as protein, though in both cases it is transcribed into RNA. So the ideas of Jennings¹ can be applicable only to some of the functions of some trinucleotides, and there must be other functions at the nucleic acid level that explain aspects of the pathology in the trinucleotide conditions. These functions presumably reflect unusual structure and/or protein-binding potential²⁻⁵ in DNA or RNA or both. While several groups have reported unusual compact structures of single-stranded CNG tracts, there is little consensus about what these may be. The safest position at the moment is that there are probably several structures, and that which one is formed depends upon length and sequence of the tract and ambient conditions⁶.

James McClellan

Sarah Newbury

Biophysics Laboratories,
University of Portsmouth,
St Michael's Building,
White Swan Road,
Portsmouth PO1 2DT, UK

1. Jennings, C. *Nature* **378**, 127 (1995).
2. Richards, R. I. et al. *Hum. molec. Genet.* **2**, 1429–1435 (1993).
3. Wang, Y.-H. et al. *Science* **265**, 669–671 (1994).
4. Yano-Yanagisawa, H. et al. *Nucleic Acids Res.* **23**, 2654–2660 (1995).
5. Taneja, K. L. et al. *J. Cell Biol.* **128**, 995–1002 (1995).
6. Mitchell, J. E. et al. *Nucleic Acids Res.* **23**, 1876–1881 (1995).

First flight

SIR — If Samuel Pierpont Langley really had successfully tested a steam-drive airplane in 1896, flying for 1.2 km before walking away from its crash¹, there would be little point in planning to celebrate the 1.0 centenary of the Wright brothers' 120-, 175- 180- and 852-foot flights of 17 December 1903 (refs 2,3). In fact, Langley's pioneering flights of 1896 were of two

26-pound unmanned machines⁴. The two launches of his manned 'Aerodrome' on the Potomac river were embarrassing failures that nearly killed his pilot and collaborator Charles Manly: the last was on 8 December 1903. Some believe that this craft (which had the backing of the Smithsonian Institution, the Congress and the US Army) could have flown had its launching catapult not failed. In retrospect, however, it seems that only the Wrights had given sufficient thought to the problems of controlling a machine in the air.

Langley died in 1906. In 1914, the Smithsonian, of which he had been an assistant secretary, allowed Glenn Curtiss to salvage Langley's machine, modify it and then succeed in a 5-second flight². The modified craft was then shamefully displayed as the world's first man-carrying aeroplane. This provoked Orville Wright eventually (in 1928) to donate the surviving Wright Flyer to the Science Museum in London. In 1942, the Smithsonian finally conceded the Wrights' priority², and received the flyer for its own collection in 1948, the year of Orville's death³.

Graham Kemp

MRC Biochemical and Clinical
Magnetic Resonance Unit,
John Radcliffe Hospital,
Oxford OX3 9DU, UK

1. Heilbron, J. L. & Bynum, W. F. *Nature* **379**, 15–18 (1996).
2. Bryson, B. *Made in America* 385–389 (Secker & Warburg/Quality Paperbacks Direct, 1994).
3. *Encyclopaedia Britannica* 15th edn. Vol. 12, p.773 (reprinted 1993).
4. *Encyclopaedia Britannica* 15th edn Vol. 7, 145 (reprinted 1993).

Not there yet

SIR — 26 October 1996 is not "the singular opportunity of celebrating the sixtieth centenary of the creation of the Universe" (J. L. Heilbron and W. F. Bynum *Nature* **379**, 15; 1996), nor is it even an option. The time span from James Ussher's creation date of 22 October 4004 BC to 22 October 1996 is just 5,999 years because, as Cesare Emiliani repeatedly pointed out in his efforts for calendar reform, there is no 'zero year' between BC and AD dates. (The time span from July of 1 BC to July of 1 AD is one year, not two). Thus all the *ab mundo condito* (AMC) anniversary dates and their corresponding "centenaries" in the article are premature by one year.

Cesare Emiliani unfortunately died last year, but how delighted he would have been at this premature celebration in *Nature*. His last communication on calendars was, in fact, published in *Nature* (as were two previous letters on the subject). Elliott and Emiliani (*Nature* **375**, 530; 1995) showed that *Nature* is in exalted company: no less an authority than Pope John Paul II proclaimed the Great Jubilee marking the start of the third millennium to be in the year 2000 rather than 2001. Both *Nature* and the Pope could very appropriately

reprise Emiliani's shortest paper (in response to a criticism of a previous paper), "Oops. Nobody's perfect". *Sic transit gloria mundi*, as Emiliani often said.

Harmon Craig

Scripps Institution of Oceanography,
University of California, San Diego,
La Jolla, California 92093, USA

SIR — Pace Heilbron and Bynum, it would surely be appropriate to commemorate the sixth millenary of "the beginning of time" according to James Ussher at the autumn equinox of 1997, that is, 1997 September 22 23.54 UTC. Some precision in the timing should be used in recognition of Ussher's own precision. In 1654, for instance, in correspondence with the Danish astronomer Nicolaus Mercator, Ussher commended a calendar employing 8 leap years in 33 years, which is indeed more accurate than the 97 leap years in 400 years of the Gregorian calendar.

In his *Annals of the Old Testament*, Ussher does not describe how he made use of astronomical tables, but he would have been aware that in 4004 BC the difference between Julian and Gregorian dates amounted to about 32 days, and thus a Julian date of 23 October 4004 BC was a Gregorian date of 21 September 4004 BC, that is, the autumnal equinox.

R. B. Elliott

77 Rathdown Park, Dublin 6, Ireland

■ J. L. Heilbron and W. F. Bynum write: Our correspondents calculate correctly within the constraints of calendrical convention. But we have written for those who, like the Pope, disdain to postpone commemorating creation out of deference to arithmetic; indeed we recommend that creation be celebrated frequently, especially every October, from now until the end. □

Please write legibly

SIR — Gill Juleff's letter about ancient iron-smelting technology in Sri Lanka (*Nature* **379**, 60–63; 1996) mentions the mediaeval crucible steel produced in India and known nowadays as *wootz*. This word has a peculiar history — at least according to the *Oxford English Dictionary*, which describes it as apparently having originated in a misprint for *wook*, representing the Canarese word *ukku*, pronounced with an initial 'w' and meaning steel. As the first example of the word is in the Royal Society's *Philosophical Transactions* of 1795 ("Dr. Scott has sent over specimens of a substance known by the name of wootz..."), it seems likely that that is when the misprint, or misreading of Dr Scott's handwriting, took place.

James Bottomley

22 Beaufort Place,
Thomson's Lane,
Cambridge CB5 8AG, UK

Giants and dwarfs meet in the middle

Alan P. Boss

The discovery of two remarkable new objects by astronomers has blurred the distinction between stars and planets. It also raises hopes of finding planetary systems like our own in the near future.

How massive can a giant planet be? What is the lowest mass for a brown dwarf star? The answers to both of these questions have changed because of observations presented at a recent meeting*. Geoffrey W. Marcy (San Francisco State University) and R. Paul Butler (University of California, Berkeley) revealed strong spectroscopic evidence for the existence of two new, low-mass companions to nearby stars. The star 47 Ursae Majoris appears to have a companion, 47 UMa B, with a mass of at least $2.3 M_J$ (Jupiter masses), while 70 Virginis is orbited by an object (70 Vir B) with a mass of at least $6.5 M_J$. The circular orbit of 47 UMa B and the moderately eccentric, star-like orbit of 70 Vir B, combined with their mass limits, imply that Marcy and Butler have found both the lowest mass star and the highest mass planet known so far, a stunning achievement indeed.

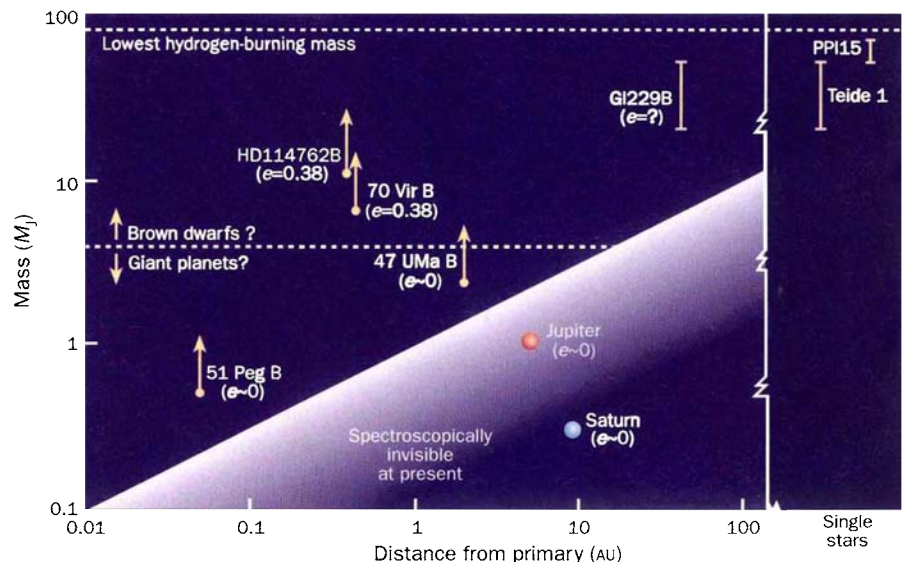
Marcy and Butler's data, which span eight years of observations, show that 47 UMa B orbits with a period of 3.02 yr at a distance of 2.1 AU (one AU is the Earth-Sun distance, about 150 million km) from 47 UMa on a nearly circular orbit (eccentricity $e = 0.0 \pm 0.15$). 70 Vir B orbits in 117 days 0.43 AU from its primary, but on an eccentric orbit ($e = 0.38$). These objects were discovered by detecting the Doppler shift produced by the radial velocity component of the primary star as it wobbles about the centre of mass of the star-companion system. Only a lower limit on the mass is found — if the companion's orbital plane is nearly face-on to us, its mass could be much larger than this lower limit. Nevertheless, for a random distribution of orbital planes, the most likely mass can be obtained by multiplying the lower limit by $4/\pi$, which results in a most probable mass of about $2.9 M_J$ for 47 UMa B and $8.3 M_J$ for 70 Vir B.

The discovery of these two new gaseous objects is noteworthy no matter what they turn out to be, but the question of whether the objects are giant planets or brown dwarfs is much more than one of semantics. The distinction is important because a brown dwarf's being companion to a nearby star says little about the likelihood that a planetary system lurks unseen within the same system. In fact, binary

stars are routinely eliminated from extrasolar planet searches, for the very good reason that we expect planets to have an easier time forming and remaining in their orbits around single stars. On the other hand, the presence of a giant planet in orbit several AU from a normal star similar to the Sun implies the existence of a retinue of lower-mass planets just waiting to be discovered. There is no doubt that 47 UMa will be thoroughly scrutinized in future searches for extrasolar planets — the race is now on to find much lower-

(Jupiter has more than three times the mass of the next largest planet, Saturn), which does not violate the theoretical prediction that there should be a sizeable gap between the mass of the biggest giant planet and the smallest brown dwarf star. The new results call that idea into question.

Planets are believed to form as a result of the collisional accumulation of rock and ice bodies called planetesimals that coagulate from interstellar dust grains orbiting in a disk around the protostar. Giant planets thereafter derive most of



Masses and separations of probable (Gl229B, 70 Vir B) and possible (HD114762B) brown dwarfs and probable giant planets (51 Peg B, 47 UMa B) in orbit around nearby stars. Shown on the right are the masses of two probable brown dwarfs (PPI15, Teide 1) which have no known companions. Along with the giant planets Jupiter and Saturn, 51 Peg B and 47 UMa B have nearly circular orbits, whereas HD114762B and 70 Vir B have eccentric ($e \sim 0.4$) orbits indicating their stellar origin. The eccentricity of Gl229B is unknown. The diagonal line roughly denotes the current limiting sensitivity of searches by the spectroscopic method.

mass, solid planets, with the 'Holy Grail' being the discovery of the first extrasolar planet capable of evolving life.

Historians of the next millennium will note that it was during the 1990s that answers to the ancient question of the existence of other planetary systems first moved from the foggy realm of theoretical speculation into the clear (albeit faint) light of hard observational fact. Last year the first planetary-mass object, 51 Peg B, was discovered orbiting a solar-type star¹.

The mass of 51 Peg B is likely to be not much more than that of Jupiter, the most massive planet in our Solar System

their mass by accretion of disk gas². The maximum mass for a giant planet was calculated³ to be about $1 M_J$, a mass at which the planet clears a gap in the disk that prevents further accretion.

Stars form in a different way, and the theoretical prediction for the minimum mass of a star⁴ is about $10 M_J$. That is the mass of the smallest clump into which a collapsing interstellar cloud can fragment during the star formation process, and is determined by the interplay of self-gravitation, thermal pressure and radiative heating during protostellar collapse. It is considerably less than the minimum mass

*187th Meeting of the American Astronomical Society, San Antonio, USA, 14-18 January 1996.

of about 80 M_J for ignition of hydrogen fusion. Stars that fall within the 10 to 80 M_J range are thus known as brown dwarfs — massive enough to have been formed in the same way as other stars, but too lightweight to shine as brightly or for very long.

The brown dwarfs that have been discovered previously all have masses that fall within the expected range of 10 to 80 M_J . PPl15 has a mass⁵ of about 60 M_J , while both Teide 1 and Gl229B have masses^{6,7} in the range 20 to 50 M_J . These masses are estimated using theoretical models of brown dwarf thermal evolution, observed luminosities and ages inferred for either the primary star or the stellar cluster they reside in. HD114762 has a companion⁸ detected in a similar way to the new objects, with a lower limit on its mass of 11 M_J , but in this case there is evidence that the orbit is viewed nearly pole-on⁹ so the mass is likely to be much greater than 11 M_J .

Thus, before the January meeting, there was a fairly wide gap between the mass of the smallest known brown dwarfs and that of the most massive planets observed or believed to be possible. The objects found by Marcy and Butler, with masses of about 3 to 8 M_J , sit right inside the gap, making it harder to use mass alone as a clear discriminator between giant planets and brown dwarfs.

However, there is another key discriminator between stellar and planetary origins: the orbital eccentricity (at least for companions at large enough distances to avoid having their orbits made circular by tidal dissipation). Normal, main-sequence binary stars are known to form with eccentric orbits¹⁰. This eccentricity is believed to be a result of the process that forms binary stars: a rapidly collapsing cloud, far from equilibrium, breaks up into two fragments moving on elliptical orbits. In contrast, major planets form with nearly circular orbits, reflecting their origin in the more ordered environment of a dissipative protoplanetary disk, which will have settled into keplerian rotation. The fact that 51 Peg B and 47 UMa B have circular orbits speaks strongly for their planetary nature, whereas the eccentricities of 70 Vir B and HD114762B imply their stellar origin.

In analogy to close binary stars, brown dwarf companions should be able to form with rather small separations, but the proximity (0.05 AU) of 51 Peg B to its star presents a problem if it formed as a planet. A giant planet cannot form closer than a few AU from its star because the icy planetesimals required for its formation are not stable in the hot inner disk¹¹. By this measure, 51 Peg B should not be there unless it is actually a brown dwarf. The problem can be solved by allowing 51 Peg B to form as a giant planet at several AU and then migrate inwards through its gravitational interactions with the disk¹². The location of 47 UMa B at 2.1 AU requires little or no orbital migration.

The spectroscopic method favours the

detection of massive companions with small separations (see figure), but patience will surely result in the discovery of extrasolar Jupiters and Saturns at greater separations than a few AU, which would imply the existence of planetary systems similar to our own. At the January meeting, Daniel Goldin (Administrator of the National Aeronautics and Space Administration) committed NASA to a long-term programme of extrasolar planet detection. NASA's future results should be even more stunning. □

Alan P. Boss is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road NW, Washington DC 20015-1305, USA.

MEDICAL GENETICS

Crystal-clear chloride channels

Steven C. Hebert

KIDNEY-STONE disease (nephrolithiasis) occurs worldwide, often causing excruciating pain and infection of the urinary tract and leading to loss of kidney function^{1,2}. The offending stones are typically composed of calcium salts that precipitate out of urine supersaturated with calcium (hypercalciuria)². This hypercalciuric nephrolithiasis is frequently inherited as a fault on the X chromosome³ and one type, Dent's disease, has been pinpointed in one family by Thakker and co-workers^{3,4} to a defect in a gene, dubbed *CLCN5*, that encodes a putative chloride channel, CLC-5.

This was quite a surprise because Dent's disease was considered to be a dysfunction in the transport of renal calcium and not apparently of chloride or fluid volume. On page 445 of this issue, Thakker and colleagues⁵ now extend their discovery of a chloride 'channelopathy' in more families with Dent's disease by identifying new mutations in the *CLCN5* gene. They also find that two other types of hypercalciuric nephrolithiasis, which map to the same defective region on the X chromosome as Dent's disease (X-linked recessive nephrolithiasis and recessive hypophosphataemic rickets), are associated with mutations in this same chloride channel which are responsible for inactivating it. This not only provides compelling evidence for an important function of the CLC-5 chloride channel in the renal tubule handling of calcium, but also poses the possibility that abnormal function of this channel may be involved in some other forms of hypercalciuric nephrolithiasis.

CLC-5 belongs to a growing family of chloride channels with twelve putative membrane-spanning helices⁶. The first member, CLC-0, was cloned from the electric organ of the *Torpedo* ray by

Jentsch and co-workers⁷. Some CLC chloride channels are widely expressed, for example CLC-2, which is involved in the regulation of cell volume. Others are tissue-specific, such as CLC-1 in muscle (mutations in CLC-1 cause dominant myotonia congenita) or the kidney-specific CLC-K1 or K2 channels, which may mediate epithelial chloride absorption. CLC-5 is most closely related, however, to the CLC-3 (rat) and CLC-4 (human) type channels, but unlike these channels, which have a broad tissue distribution, both rat⁸ and human³ CLC-5 channels are expressed predominantly in kidney.

A central question now is how loss of the function of this renal chloride channel can result in hypercalciuria, as well as in the other characteristic tubule transport defects. In the nephron, the functional unit of the kidney, ionized calcium is freely filtered out of the blood plasma, along with waste products destined for excretion, at the glomerulus and into the proximal tubule, where it is reabsorbed as necessary or otherwise passed into the distal tubule for excretion. This carefully controlled process goes wrong in hypercalciuria as a result of defective reabsorption of the divalent cation.

Although no information is currently available as to which specific epithelial segments of the nephron express CLC-5 channels, the presence of small proteins in the urine of patients with Dent's disease points to some problem with the tubules, probably involving, at a minimum, the proximal tubule segments. Most proteins are too large to be filtered at the glomerulus, but small proteins like β_2 -microglobulin can get through and are normally reabsorbed in this region of the nephron. Other indicators of damage to the proximal tubules (for example, the presence of

1. Mayor, M. & Queloz, D. *Nature* **378**, 355–359 (1995).
2. Pollack, J. B. A. *Rev. Astr. Astrophys.* **22**, 389–424 (1984).
3. Lin, D. N. C. & Papaloizou, J. *Mon. Not. R. astr. Soc.* **191**, 37–48 (1980).
4. Boss, A. P. *Astrophys. J.* **331**, 370–376 (1988).
5. Stauffer, J. R., Hamilton, D. & Probst, R. G. *Astr. J.* **108**, 155–159 (1994).
6. Rebolo, R., Zapatero Osorio, M. R. & Martín, E. L. *Nature* **377**, 129–131 (1995).
7. Nakajima, T. et al. *Nature* **378**, 463–465 (1995).
8. Latham, D. W. et al. *Nature* **339**, 38–40 (1989).
9. Cochran, W. D., Hatzes, A. P. & Hancock, T. J. *Astrophys. J.* **380**, L35–L38 (1991).
10. Duquennoy, A. & Mayor, M. *Astr. Astrophys.* **248**, 485–524 (1991).
11. Boss, A. P. *Science* **267**, 360–362 (1995).
12. Lin, D. N. C. & Bodenheimer, P. *Nature* (submitted).

amino acids, uric acid or glucose in the urine) are also found in some patients with X-linked recessive nephrolithiasis⁹.

About 70–80 per cent of the filtered calcium is reabsorbed in the proximal tubule in a largely passive process that is dependent on solvent drag associated with bulk fluid reabsorption¹⁰, which in turn is dependent on reabsorption of active solute, including chloride. One possibility is that CLC-5 channels in the proximal tubule are involved in transepithelial chloride transport; loss of this channel would result in diminished reabsorption of chloride and fluid and consequently of calcium. But how such a mechanism could account for the presence of low-molecular-weight proteins in the urine is not clear.

CLC-5 chloride channels, like a number of other types of ion channels, appear to form multimeric complexes^{6,11} which might not only influence the biophysical properties of CLC-5-containing complexes, as suggested by Thakker and colleagues, but could also permit the association of CLC-5 with other subunits, expanding the repertoire of possible physiological roles for this channel. For example, low-molecular-weight proteins that are specifically absorbed in the proximal tubule are taken up by endocytosis into a vacuolar-lysosomal system with an acidic interior provided by the action of an electrogenic H⁺-ATPase pump¹². Chloride channels are important in such vesicular systems because they provide a mechanism for dissipating the charge that results from proton pumping.

Perhaps CLC-5 channels play some part in this process. Could the disrupted movement of these vesicles within the cell affect the function of the transporters for calcium (and maybe those for phosphate or amino acids) in renal tubule plasma membranes, and therefore the absorption of these solutes? Or could CLC-5 channels also be involved in the control of calcium transport in more distal segments of the human nephron? Regulated calcium absorption occurs in these more distal nephron segments and divalent mineral ion absorption is strongly dependent upon, or influenced by, chloride transport processes¹⁰. The possible participation of CLC-5 in these processes needs to be explored.

Another big question to emerge from the new work⁵ concerns the differences found among these three hereditary hypercalciuric disorders in bone pathol-

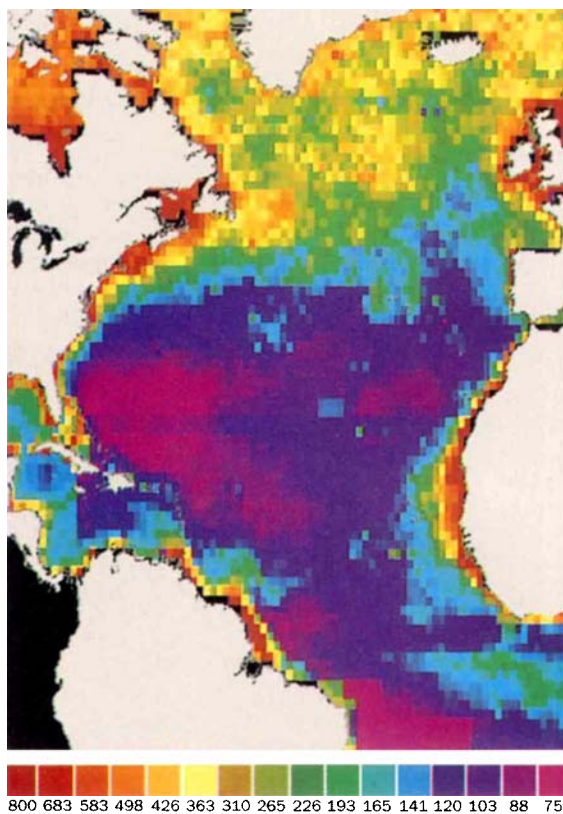
Provincial primary production

Is this picture just another ocean-colour satellite image of the North Atlantic? Not at all. S. Sathyendranath *et al.* have exploited the remote-sensing revolution — which transformed marine science with ocean-colour measurements made using a trial sensor between 1978 and 1986 — by calculating the biological productivity of the entire North Atlantic at a regional resolution and by season (*Deep-Sea Res. I*, 42, 1773–1802; 1995). The picture shows the estimated annually averaged primary production in 1979, the year with the best total data set (units are g carbon m⁻²).

The authors went about their task by first devising a scheme for dividing the ocean up into 18 different 'biogeochemical provinces' according to observed algal ecology and physical processes such as vertical mixing — these provinces are akin to different terrestrial biomes, boreal forest, savanna, desert and tundra for instance. They then used thousands of *in situ* measurements of photosynthetic parameters and vertical distributions of chlorophyll, combined with phytoplankton pigment data derived from the satellite observations, to compute primary production for different seasons for each biogeochemical province. Despite the patchiness of the satellite data, these are the most robust regionally resolved estimates yet made of marine primary productivity on a basin-wide scale. Such estimates are of great value not only to ecophysiologicalists, but — given the involvement of marine ecosystems in the global carbon cycle — also to climate researchers and oceanographers.

Plans for the deployment of advanced purpose-built sensors have been frustrated for the past few years. But eleven such sensors are scheduled for launch before the year 2000, several of them by the end of 1996. They will carry with them hopes that the ensuing data will ultimately provide estimates of the rates of other key processes within the global carbon cycle, such as new production and air-sea exchange of carbon dioxide.

Philip Newton



1. Johnson, C. M., Wilson, D. M., O'Fallon, W. M., Malek, R. S. & Kurland, L. T. *Kidney Int.* **16**, 624–631 (1979).
2. Coe, F. L., Parks, J. H. & Asplin, J. R. *New Engl. J. Med.* **327**, 1141–1152 (1992).
3. Fisher, S. E. *et al. Human molec. Genet.* **3**, 2053–2059 (1994).
4. Fisher, S. E. *et al. Genomics* **29**, 598–606 (1995).
5. Lloyd, S. E. *et al. Nature* **379**, 445–449 (1996).
6. Steinmeyer, K., Ortland, C. & Jentsch, T. J. *Nature* **354**, 301–304 (1991).

7. Jentsch, T. J., Steinmeyer, K. & Schwarz, G. *Nature* **348**, 510–514 (1990).
8. Sakamoto, H., Kawasaki, M., Uchida, S., Sasaki, S. & Marumo, F. J. *Am. Soc. Nephrol.* **6** (abstr.), 350 (1995).
9. Frymoyer, P. A. *et al. New Engl. J. Med.* **325**, 681–686 (1991).
10. Friedman, P. A. & Gesek, F. A. *Physiol. Rev.* **75**, 429–471 (1995).
11. Steinmeyer, K., Lorenz, C., Pusch, M., Koch, M. C. & Jentsch, T. J. *EMBO J.* **13**, 737–743 (1994).
12. Christensen, E. I. & Nielsen, S. *Semin. Nephrol.* **11**, 414–420 (1991).

tribute to bone abnormalities? Perhaps other genetic or environmental factors, such as calcium or phosphate intake, dietary vitamin D, or circulating levels of parathyroid hormone, may influence the severity of bone and renal pathology in these patients.

Once the localization, function and regulation of CLC-5 are known, we should have a clearer picture of how this chloride channel functions in the kidney. But there may be more surprises in store.

Steven C. Hebert is in the Laboratory of Molecular Physiology and Biophysics, Renal Division, Department of Medicine, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA.

Dangerous liaisons and disease

David W. Macdonald

CARNIVORES, often made rare by people's actions, may be annihilated by diseases, often caught from people's pets. Can biologists help? Lion researchers, writing on page 441 of this issue¹, show that the answer can be 'yes'. The group concerned — Roelke-Parker *et al.* — demonstrate that the cause of death of an estimated 1,000 lions in the Serengeti National Park during 1994 was canine distemper virus (CDV), probably transmitted from domestic dogs. The disease then spread northwards from Tanzania to the Maasai Mara reserve in Kenya.

Canine distemper virus, like measles a morbillivirus, is notorious in conservation circles. One variant almost drove black-footed ferrets, *Mustela nigripes*, to extinction in the 1980s (ref. 2), and another exploded among seals and dolphins a few years later (ref. 3). Now we have the case of the Serengeti lions, which raises a whole raft of issues in conservation biology — not least the possible transmission of viruses in particular from domestic animals to endangered species, and how such outbreaks of disease might be prevented.

No mammal, ourselves included, can be complacent in the evolutionary battle against pathogens. Roelke-Parker and colleagues also found that blood sera hoarded from long-dead Serengeti lions contain antibodies against CDV, proclaiming that they had survived earlier encounters with the virus. But close to the driveshaft of the evolutionary engine toils the notorious Red Queen, who runs perpetually just to hold her own. As pathogens race to outwit our immune systems they gain a massive head start through the brevity of their generations — an advantage made memorable by knowing that the bacteria in your gut will pass through six times as many generations in your lifetime as people have since they were apes⁴. Just as our own Red Queen apparently lost races with green monkey disease and the human immunodeficiency virus, so it seems that the Serengeti lions, having resisted the CDV of the 1980s, have succumbed to the 1990s version. Aside from the practical point that the people who had squirrelled away historical serum samples have amply illustrated the value of long-term screening of wildlife diseases, the main message to emerge from this shift in susceptibility is that prospects for endangered populations may quickly take a turn for the worse.

Such threats are alarmingly evident

from the plights of several rare carnivores⁵. The Ethiopian wolf, *Canis simensis* — the world's rarest canid, found only on a few Ethiopian plateaux — was, by the late 1980s, reduced to a mere 400 individuals in its main stronghold in the Bale Mountains. Rabies, probably contracted from village dogs, swept through this population in 1991. Some two years later more than half of the Bale Mountain wolves were dead⁶. Similarly, the tiny Blanford's fox, *Vulpes cana*, a dainty pow-

der-puff of a creature whose naked footpads and flamboyant tail adapt it to bouncing improbably up sheer cliffs, had scarcely been discovered in Israel when its existence was jeopardized by a rabies epidemic, probably ignited in domestic dogs and fanned by populous red foxes, *Vulpes vulpes*.

Other examples are genealogically closer to the Serengeti lions. These include the much debated spectre of diseases, such as feline infectious peritonitis, in the genetically homogeneous cheetah, *Acinonyx jubatus*⁷, and the threat of the transmission of feline immunodeficiency virus (FIV) from domestic cats to the Florida panther⁸, *Felis concolor*, or to the 80 or so surviving Iriomote cats, *F. iriomotensis*, that have no antibodies against it⁹. Scottish wildcats, *F. sylvestris*, already imperilled by genetic introgression from their domestic descendants may also be threatened by FIV from farm cats¹⁰.

Dangerous liaisons need not be directly with domestic species. The outbreak of CDV in 1994, which accounted for a third of the lions in the Serengeti, may not threaten their population (although it will

doubtless have reverberations throughout the ecosystem). But it could have made them fatal companions: had the few remaining Serengeti wild dogs, *Lycaon pictus*, not already disappeared in 1992, lion distemper might have finished them off. As it happens, following a long decline, some Serengeti wild dogs, representatives of the second most endangered of all canids, succumbed to rabies, a disease of which the canine symptoms are notoriously similar to those of CDV.

So what can we do in such cases? Two things, at least. First, tackle the background risk by reducing the numbers of infectious domestic mammals (rinderpest, a morbillivirus that affects antelope, was controlled in East Africa by vaccinating cattle and goats). Second, concentrate on the imperilled wildlife, perhaps by vaccinating them. This is an approach that triumphed in the control of fox-borne rabies in Europe⁵. In the case of the Serengeti wild dogs, vaccination was attempted¹¹. As it happens it did not work — perhaps, as some maintain (see ref. 12 and compare with ref. 13), because the attempt was flawed; or perhaps, as others suspect¹⁴, because the animals immunized against rabies were killed by another pathogen such as distemper. The answer to that debate is lost in history. Meanwhile the 4 per cent annual growth rate of the human population (and its dogs) around the park adds urgency to the quest for practical solutions.

Molecular sleuthing pointed the finger at these domestic dogs as the source of the risk to the Serengeti lions. Some might consider them over-populous, useless curs, but others, closer at hand, give them names and fondle them. Should the dogs be left there, a virus factory fulminating on the borders of a unique ecosystem? Clearly not, but equally clearly there is an ethical impropriety, and probably also political and logistical impossibilities, of removing them.

In a sequel to Roelke-Parker and colleagues' work, Project Life Lion, working with the Tanzanian Veterinary Service, has launched a dog-vaccination campaign. Rather than fretting over whether rabies or distemper poses the greater threat to wildlife, they are vaccinating against both. This is a campaign which I gather delights the villagers, who naturally prefer their

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Steven C. Kaufman/Bruce Coleman

1. Roelke-Parker, M. E. *et al.* *Nature* **379**, 441–445 (1996).
2. Williams, E. S. *et al.* *J. Wildl. Dis.* **24**, 385–398 (1988).
3. Osterhaus, A. D. M. E. *et al.* *Nature* **335**, 403–404 (1988).
4. Ridley, M. *The Red Queen* (Viking, London, 1993).
5. Macdonald, D. W. *Onderstepoort J. Vet. Res.* **60**, 351–355 (1993).
6. Sillero-Zubiri, C. *et al.* *J. Wildl. Dis.* **32**, 80–86 (1996).
7. Heeney, J. L. *et al.* *J. Virol.* **64**, 1964–1972 (1990).
8. Roelke, M. E. *J. Wildl. Dis.* **29**, 36–49 (1993).
9. Mochizuki, M. *et al.* *J. Wildl. Dis.* **26**, 236–245 (1990).
10. Yamaguchi, N. *et al.* *Epidemiol. Infect.* (in the press).
11. Gascoigne, S. *et al.* *J. Wildl. Dis.* **29**, 396–402 (1993).
12. Burrows, R. *Nature* **359**, 277 (1992).
13. Ginsberg, J. *et al.* *Conserv. Biol.* **9**, 665–674 (1995).
14. Macdonald, D. W. *et al.* *Nature* **360**, 633–634 (1992).

dogs in good health; which may save imperilled wildlife; which uses First World money to protect an ecosystem of global importance; and which has grown out of a quest to consider seemingly arcane questions about animal populations and molecular evolution, thereby showing how good science is an indispensable element of good conservation.

Faced with diseases, small populations

of imperilled species are extinctions waiting to happen. Roelke-Parker *et al.* show how ecological research can both untangle the problem and provide the impetus for solving it. □

David W. Macdonald is in the Wildlife Conservation Research Unit, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

ASTRONOMY

X-rays reveal heart of darkness

Michael J. West

IN Antoine de Saint Exupéry's story *The Little Prince*, a fox reveals an important secret about the world: "What is essential is invisible to the eye". This same conviction is shared by astronomers, who for over half a century have believed that as much as 90 per cent of all the matter in the Universe is in an invisible form, which can be detected only by its gravitational influence on the luminous matter we can see, such as stars and galaxies¹. Yet what this dark matter is composed of and how it is distributed throughout space are still not known.

On page 427 of this issue, Ikebe *et al.*² report on the distribution of dark matter in one nearby region of the Universe, the Fornax cluster of galaxies (see figure). Based on measurements made with ASCA (Advanced Satellite for Cosmology and Astrophysics), a Japanese X-ray satellite that was launched in February 1993, they have found evidence that the dark matter there is clustered on at least two distinct scales. They argue that this points to a hierarchical distribution of dark matter, or might be indicative of more than one type.

The existence of dark matter in clusters of galaxies was first suggested³ by Fritz Zwicky in 1933. Zwicky's seminal study of the Coma cluster, an immense system of thousands of galaxies, showed that the total mass needed to gravitationally bind this cluster exceeds the amount of matter that is present in the visible galaxies by roughly an order of magnitude. Three years later, Sinclair Smith showed⁴ that a similar discrepancy exists between the gravitationally inferred mass of the Virgo cluster and the sum of the masses of its constituent galaxies. Thus was born one of the great mysteries of modern astrophysics, which has come to be known as the 'dark matter problem'.

Dynamical evidence of dark matter has since been found in many other astronomical systems, from individual galaxies (including our own Milky Way) to the largest observed structures in the Universe. Clearly, then, any attempt to understand the origin and evolution of all these

objects requires some knowledge of the distribution and properties of the apparently ubiquitous dark matter. It seems unlikely that it can all be made of the familiar 'baryonic matter' of atoms and ions. The primordial abundance of different elements in the Universe, as well as the isotropy of the microwave background radiation that is a relic of the Big Bang,



The central region of the Fornax cluster of galaxies, where X-ray observations have revealed two distinct distributions of dark matter.

all but rule out that possibility. So astronomers are searching for a completely new form of matter.

But how does one study dark matter if one cannot see it? One way is to look at the X-ray emissions from clusters of galaxies. Besides their luminous galaxy populations, many clusters also contain diffuse, hot gas at temperatures of 10^7 to 10^8 K, which emits X-rays by thermal bremsstrahlung (the 'braking radiation' emitted by an electron accelerating rapidly in the electric field of a positive ion). This gas is likely to be in hydrostatic equilibrium, where the thermal pressure of the hot gas is balanced by the gravity of the cluster. Regions of higher density will trap more of the gas, so the X-ray emission should trace the distribution of

dark matter, which dominates the gravitational field.

It is this method that Ikebe *et al.* have used to map the dark matter content of the Fornax cluster of galaxies, a relatively nearby example which has the supergiant elliptical galaxy NGC1399 at its centre. By modelling the X-ray gas emission measured by ASCA, they have shown quite convincingly that the dark matter distribution in Fornax is a mixture of two distinct components on different scales, one associated with the cluster as a whole and the other, smaller one belonging to the giant galaxy.

The authors claim that their results "provide unambiguous observational evidence for the hierarchical nature of the dark matter distribution". Previous X-ray observations and studies of gravitational lensing (the gravitational bending of light rays to form cosmic 'mirages'), had clearly established the existence of dark substructures within clusters of galaxies⁵⁻⁹. Nevertheless, Ikebe *et al.* do provide strong additional support for the notion that the dark matter content of the Universe is arranged in a continuous hierarchy of structures from small to large scales. Such a hierarchy is a well established characteristic of the distribution of visible matter, with stars grouped into galaxies, galaxies assembled into clusters, and clusters embedded in larger superclusters.

The new results will undoubtedly be reassuring to many cosmologists who have long assumed, with little hard evidence, that the invisible matter has a similar hierarchical distribution to that of the luminous material. This hypothesis lies at the heart of most currently popular models for structure formation in the Universe, which propose that galaxies and larger structures originated from small clumps of dark matter that clustered together to form progressively larger objects.

An interesting question raised by the authors is whether the apparent clustering of dark matter on two different scales in Fornax might indicate the presence of

1. Tremaine, S. *Physics Today* **45**, 28-36 (1992).
2. Ikebe, Y. *et al.* *Nature* **379**, 427-429 (1996).
3. Zwicky, F. *Helv. phys. Acta* **6**, 110-132 (1933).
4. Smith, S. *Astr. J.* **83**, 23-30 (1936).
5. Jones, C. & Forman, W. in *Clusters and Superclusters of Galaxies* (ed. Fabian, A. C.) 49-70 (Kluwer, Dordrecht, 1992).
6. White, S. D. M., Briel, U. G. & Henry, J. P. *Mon. Not. R. astr. Soc.* **261**, L6-L12 (1993).
7. Slezak, E., Durret, F. & Gèrald, D. *Astr. J.* **108**, 1996-2008 (1994).
8. Pelló, R., Le Borgne, J.-F., Soucail, G., Mellier, Y. & Sanahuja, B. *Astrophys. J.* **366**, 405-411 (1991).
9. Kneib, J.-P., Mellier, Y., Fort, B. & Mathez, G. *Astr. Astrophys.* **273**, 367-376 (1993).
10. Klypin, A., Holtzman, J., Primack, J. & Regos, E. *Astrophys. J.* **416**, 1-16 (1993).

more than one type of material. Such speculation is not unfounded. So-called 'mixed dark matter' models, which postulate the existence of different types of non-baryonic dark matter particle, have come into vogue in recent years as cosmologists scramble to reconcile much-cherished theories with ever more stringent observational constraints. Particularly successful have been hybrid models based on a mixture of 70–80 per cent 'cold' dark matter (CDM) and 20–30 per cent 'hot' dark matter (HDM), plus a smattering of baryons which are responsible for producing the light that we see¹⁰.

The favourite CDM candidates include 'axions' and other exotica predicted by some particle physics theories, while the most likely HDM constituent would be one or more species of massive neutrino,

if neutrinos do indeed have mass. If the mixed model is correct, then galaxy dark matter haloes might plausibly be composed mainly of CDM, while clusters could contain a greater fraction of HDM. Sceptics might counter, however, that it is premature to conjecture the existence of two forms of non-baryonic dark matter when the existence of even one type has not yet been established. If direct experimental evidence of any CDM or HDM particle is obtained in the lab, cosmologists will then be a giant step closer to understanding structure formation in the Universe. □

Michael J. West is in the Department of Astronomy and Physics, Saint Mary's University, Halifax, Nova Scotia B3H 3C3, Canada.

GENE EXPRESSION

When the junk isn't junk

Melissa J. Moore

EVER since their discovery almost 20 years ago^{1,2}, introns have been much maligned as examples of 'junk DNA', a catch-all phrase for chromosomal sequences with no apparent function. Although this view has changed somewhat, there is still a tendency to ignore intronic sequences in favour of the so-called 'exonic' ones when trying to establish the function of a newly identified gene. Tycowski and colleagues' description of a gene in which all the functional sequences are seemingly contained within

maintaining both proteins and stable RNA components. The latter RNAs (ribosomal or rRNAs) are examples of the other possible fate for an RNA transcript: one in which the RNA itself becomes the final and functional gene product.

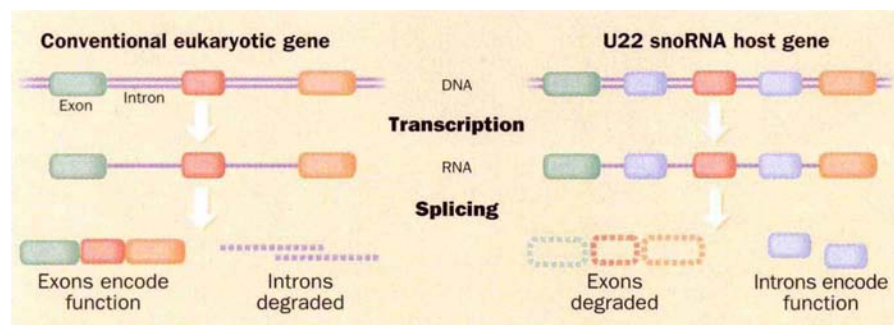
In bacteria, the genetic information in DNA is collinear with the specified product, so newly transcribed mRNAs can be used directly as templates for translation. In eukaryotes (organisms with nucleated cells), however, genes are not always contiguous, but are often interrupted by

cle⁴ of 1978, in which the terms 'intron' and 'exon' were coined, Walter Gilbert suggested that "The gene is a mosaic: expressed sequences held in a matrix of silent DNA, an intronic matrix". According to this view, the principal but passive role of introns is to speed the genetic recombination (shuffling) of exonic sequences, greatly expediting the rate at which new protein-coding genes can evolve. That introns might have active functions in their own right was not fully appreciated until the remarkable discovery in 1982 of self-splicing introns⁵. When transcribed, these RNA sequences fold into highly ordered tertiary structures that can catalyse self-removal and re-ligation of the host exons without the aid of a spliceosome. Interestingly, many of these self-splicing introns double as mRNAs from which proteins are expressed that assist self-splicing and/or mediate intron translocation to new DNA sites⁶. So introns can also contain expressed regions.

Another kind of functional RNA occasionally found lurking within introns of genes whose exons specify proteins involved in translation are snoRNAs, small, stable RNAs that accumulate in the nucleolus⁷. The nucleolus is a large substructure of the nucleus dedicated to rRNA transcription and processing and subsequent ribosome assembly. Although their exact functions are not yet known, several snoRNAs are required for proper rRNA processing. Intronic snoRNAs are transcribed as part of the parent pre-mRNA. After intron excision, exonucleases trim back the surrounding intron sequences to yield the mature snoRNA⁸.

In 1994, Tycowski *et al.*⁹ initially characterized human U22 snoRNA (which is required for processing the rRNA of the smaller ribosomal subunit), and determined a partial sequence for the U22 host gene (UHG). The U22 sequence is embedded in the penultimate intron of the UHG, but unlike other genes that host intronic snoRNAs, the UHG mRNA did not seem to encode any protein. Because functional regions of genes are usually conserved between species, the authors undertook a second cloning project to sequence the mouse UHG, and they now report a comparison of the complete human and mouse sequences³.

As was suspected, it turned out that no protein-coding regions are conserved between the two UHG mRNAs. But comparison of the intron sequences revealed a big surprise: the group had hit the snoRNA jackpot. Instead of just U22, UHG actually encodes eight snoRNAs, one each in eight out of its nine (mouse) or ten (human) introns. The seven other snoRNAs, U25 to U31, had not been described before, so the authors examined their disposition. All seven are stably expressed and complexed with fibrillarin, a nucleolar protein antigen



Comparison of exon and intron fates in a conventional eukaryotic gene (left) versus the U22 snoRNA host gene (right).

the introns, and not the exons (page 464 of this issue³), should, however, give us pause to reconsider this prejudice.

Genes are segments of DNA that encode functional biopolymers, either RNA or protein. The first step in gene expression is transcription of a full-length RNA copy. In most cases these RNA transcripts then serve as the templates or messengers (mRNAs) for protein synthesis. This translation of information from mRNA to protein is mediated by ribosomes, macromolecular machines con-

sequences not represented in the final biopolymer. Consequently, most protein-coding transcripts must be processed to remove these intragenic regions or 'introns' before protein expression can occur (see figure). The process by which introns are removed and the flanking 'exons' (expressed regions) stitched back together, called pre-mRNA splicing, is carried out by spliceosomes, biochemical machines of similar size and complexity to ribosomes.

In his widely cited News and Views arti-

commonly associated with functional snoRNAs. Tycowski *et al.* also discovered the fate of the UHG mRNA: it is rapidly degraded once it reaches the cytoplasm. Thus it would seem that the UHG is an example of a totally 'inside-out' gene in which only the introns encode function, whereas the spliced exons are just left-over junk to be discarded (see figure).

With elucidation of the UHG structure, it is clear that the original definition of an exon as an expressed region no longer holds. Today the terms intron and exon simply signify RNA sequences that become physically separated during the process of RNA splicing. Conclusions as to which parts of a gene express function must be based on other analyses.

Questions of immediate interest regarding the UHG are how and why this gene with such a remarkable structure came to be. The 'how' is probably through intron shuffling, which invokes the same recombinational mechanisms originally proposed for exon shuffling⁴. Evidence of a recent intron-shuffling event is seen in the U22-like sequence in the extra intron of human UHG. As for the 'why', Tycowski *et al.*³ report that the UHG promoter resembles those of the ribosomal-protein genes. Perhaps the UHG exons originally encoded a protein whose function was subsequently obviated by one or more of its internally encoded snoRNAs. Because of this promoter, UHG transcription is probably coordinated with that of other components required for ribosome biogenesis. If their functions mean that U22 and U25–U31 snoRNAs must always be present at equimolar ratios, encoding them within a single transcript is quite an efficient way to ensure that happens.

Is the UHG structure unique? Probably not. Given the variety of genome-sequencing projects now well underway, interspecies comparisons will inevitably reveal innumerable conserved intronic sequences. Some percentage of such introns will surely be flanked by non-coding exons, which will add to a growing heap of exonic 'junk DNA'. After all, turn-about is only fair play. □

Melissa J. Moore is in the Department of Biochemistry, Brandeis University, 415 South Street, Waltham, Massachusetts 02254, USA.

- Berget, S. M., Moore, C. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3171–3175 (1977).
- Chow, L. T., Gelinas, R. E., Broker, T. R. & Roberts, R. J. *Cell* **12**, 1–8 (1977).
- Tycowski, K. T., Shu, M.-D. & Steitz, J. A. *Nature* **379**, 464–466 (1996).
- Gilbert, W. *Nature* **271**, 501 (1978).
- Kruger, K. *et al. Cell* **31**, 147–157 (1982).
- Lambowitz, A. M. & Belfort, M. A. *Rev. Biochem.* **62**, 587–622 (1993).
- Maxwell, E. S. & Fournier, M. J. A. *Rev. Biochem.* **64**, 897–934 (1995).
- Kiss, T. & Filipowicz, W. *Genes Dev.* **9**, 1411–1424 (1995).
- Tycowski, K. T., Shu, M.-D. & Steitz, J. A. *Science* **266**, 1558–1561 (1994).

Illuminating entanglement

Tony Sudbery

RICHARD Feynman's elaboration of the two-slit experiment¹ is a classic discussion of wave-particle duality and the uncertainty principle. A colleague and I use it as a script for a knockabout dialogue on quantum mechanics for a general audience, but we have to admit, if pressed, that the experimental results with which we hit each other are purely imaginary. This, we say, is what would happen if anyone were clever enough to do the experiments. Now Michael Chapman and collaborators at MIT have brought Feynman's sequence of ideal demonstrations closer to realization in the laboratory².

Part of Feynman's dialogue follows. A professor is explaining the restrictions placed by quantum mechanics on what we can know about the movement of particles at the microscopic level. He shows the interference pattern made on a screen when particles travel to it through a wall with two open slits (Fig. 1a), and explains that this could not happen if each particle went through just one of the two slits.

Professor: ...so we must abandon the common-sense idea that each atom passes through just one of the slits.

Student: Hey, wait a moment. You don't expect us to accept that, do you? You've just shown us that these things are particles: they hit the screen at a sharp point. So they must have gone through one of the slits. You can prove it by

shining light on the slits; it will be reflected off the particle as it comes through, so you can see which slit it is at.

Professor: Yes, that's true. But don't forget that the light will have an effect on the particle; it will change its momentum unpredictably, and this destroys the interference pattern (Fig. 1b).

Student: Well, don't do such a clumsy experiment, then. Use a less intense beam of light, so that it will have less effect on the particles.

Professor: Now you're forgetting the quantum nature of light. A beam of light consists of photons, each with a definite energy (remember $E=h\nu$). Reducing the intensity of the light only reduces the number of photons, and therefore reduces the chance that the atom will be seen. Each photon that hits an atom will have the same effect on it as before, and for these atoms the interference pattern will disappear. For those atoms that are missed by the photons, the interference pattern remains; but we can't see which slit they've come through.

Student: OK, then reduce the effect of each photon. Use light with a lower frequency.

Professor: You're right, that will have less impact on the atoms; and if the frequency is low enough the impact will be small enough not to destroy the interference pattern. But low-frequency light has a long wavelength, and can't resolve short distances; so you can't use it to distinguish the two slits. I'm afraid you can't get round it; when the atoms are showing interference effects, they cannot be regarded as particles. ▶

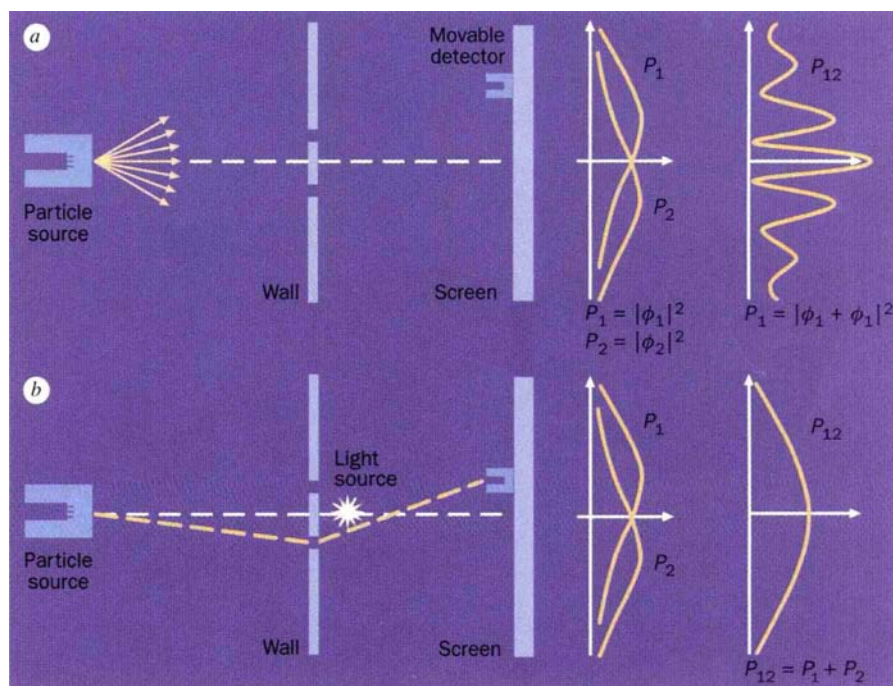


FIG. 1 Patterns made by particles hitting a screen in the two-slit experiment. a, When they are not observed they behave as coherent waves, and their probability distribution P_{12} is the result of adding wave amplitudes ϕ_1 and ϕ_2 due to each slit; in other words, they form an interference pattern. b, The interference pattern is destroyed by using a light source to find out which path a particle takes; now the particles behave like classical bullets, and their probability distribution is just that obtained by adding the single-slit distributions. (Figure adapted from ref. 1.)

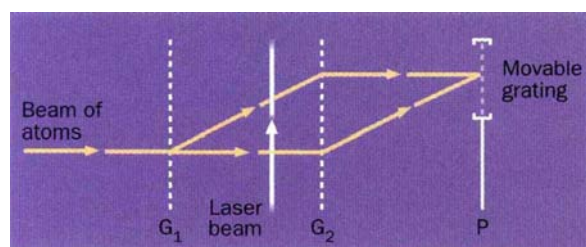


FIG. 2 The experiment uses an interferometer consisting of two parallel diffraction gratings G_1 and G_2 . There are two different ways that atoms can move from the first grating to a plane P beyond the second grating. In this plane there is a pattern of interference between the two beams. As in the two-slit experiment, this interference pattern would not be possible if each atom took one definite route through the apparatus. It is possible to observe the atoms by shining a laser beam on them as they pass between the first two gratings, and focusing the scattered light to show which path the atom was on. Chapman *et al.* show that the pattern is indeed destroyed, even though in their experiment the light was not actually focused: it is enough that it could have been. (Figure adapted from ref. 2.)

The experiment of Chapman *et al.* does not have two slits, but it retains the essential feature that the beam of atoms is split into two parts that move along different paths and show an interference pattern when they are recombined, the interference being lost when the atoms are observed by means of a laser beam to see which way they went (Fig. 2). The third stage of the dialogue, the effect of reducing the frequency of the light until its wavelength is greater than the separation between the two possible paths, was realized in a different way: the frequency of the light was kept the same, but the path separation was changed by moving the laser beam. As predicted by Feynman, the interference is present when the wavelength of the light is longer than the separation between the paths, and is progressively lost as the separation increases.

But there is a surprise: the interference is restored again as the separation increases still further. This can be seen as an effect of optical diffraction. A single photon of wavelength λ can give information on which of two point objects scattered it if the separation of the objects is $\lambda/2$, but not if the separation is $3\lambda/4$, because we cannot then tell whether the photon is arriving at the screen in the centre of the image of one of the objects or in the first diffraction ring surrounding the image of the other object.

Even when the atoms are observed when the two paths are well separated, so that the scattered light potentially contains definite information about the path, it is still possible to restore the interference pattern. In order to collect the information the light must be focused, which makes it impossible to know its direction of travel. Conversely, measuring the direction of the scattered photon destroys the information about the path of the atom. Chapman's group modified their experiment so as to do something equivalent to

this, and found that those atoms which scattered the photons in any particular direction fell into an interference pattern. The patterns associated with different photon directions had different phases, so that the overall appearance was of no interference.

This illustrates a general feature of quantum measurement processes. Before a measurement is made on a system, several different potential results coexist in a delicate balance; the measurement affects the system by apparently destroying this balance, or 'coherence'. In the two-slit experiment,

observing the path of the particle destroys the coherence which makes the interference pattern possible. However, the coherence is not really lost: it has been extended to involve properties of the measuring apparatus — the apparatus and the system are said to be 'entangled'. If the apparatus is large and complicated, its state cannot be known in sufficient detail to reveal this coherence. In the experiment of Chapman *et al.*, however, the apparatus could be taken to be the single photon that scatters off the atom, and the coherence in the extended system consisting of the photon and the atom is revealed in the interference patterns associated with photons moving in particular directions. In general, one can always delay the loss of coherence until one has made a further observation on the measuring system. This is von Neumann's principle of the arbitrariness of the boundary between the observed system and the observing apparatus.

A slight extension of the experiment pushes one uncomfortably close to paradox. In principle one could delay the observation of the photons, and the decision whether to focus them or measure their direction, until after the atoms had formed their pattern. Then one and the same pattern could be interpreted either as the sum of two non-interference patterns formed by atoms which went along the two different paths, or as the sum of many interference patterns formed by atoms which could not be said to have gone along either of the paths. Einstein would not have been happy. □

Tony Sudbery is in the Department of Mathematics, University of York, York YO1 5DD, UK.

1. Feynman, R. P., Leighton, R. B. & Sands, M. L. *The Feynman Lectures on Physics* Vol. 3, 1-4 (Addison-Wesley, Reading, 1965).
2. Chapman, M. S. *et al.* *Phys. Rev. Lett.* **75**, 3783-3787 (1995).

DAEDALUS

Smooth reactions

HEAT, the most fundamental tool of chemistry, is a very blunt instrument. It accelerates all reactions, particularly useless and destructive ones. Even with the aid of catalysis, the oxidation of ammonia to nitric oxide is always liable to go to valueless nitrogen, while that of ethylene to ethylene oxide may run away to carbon dioxide. And who knows what you might get by heating gunpowder if it didn't explode first?

Heating, of course, is simply an uncontrolled bombardment by thermal phonons of all energies. Last week Daedalus devised a novel monochromatic 'tuned heat' whose phonons all had the same energy. He generated it by hitting the sample with a uniform electron-beam of precisely defined energy. He now argues that tuned heat should be ideal for promoting specific chemical reactions. It could give the molecules the precise activation energy needed for one desired reaction, while leaving side-reactions unexcited. DREADCO's chemists are trying it. They are beaming monochromatic phonons into a wide variety of reaction mixtures. By accurate tuning, they should be able to promote just one reaction out of many. Competing reactions, including that bugbear of all chemistry, general thermal decomposition, will be left out in the cold. With their competitors frozen out, even previously impossible reactions could be tickled along nicely.

One reaction crying out for this treatment is depolymerization. Even pure polymers cannot usually be turned back to the original monomer. A mixture of plastic waste seems even more hopeless. But properly tuned heat could activate one depolymerization at a time. All the polystyrene (say) in the mixture could be converted neatly to styrene, and distilled off to be used afresh. A quick change of tuning could then unzip the PVC to vinyl chloride, then the polythene to ethylene gas, and so on in sequence. Other waste materials could be reclaimed by selectively tuned hydrolysis. Wood waste could go to glucose and lignin monomers, and motor oil to alcohol.

Tuned heat should also tame that jungle of thermal decomposition, cooking. Normal heat promotes pleasant and acrid flavours, browning and charring, tenderizing and toughening, all at once. High skill is needed to get the best from the raw material. But the DREADCO tuned oven will coax entrancing flavours, textures and shades from the cheapest and most unpromising ingredients — even vegetarian. Hopeless cooks everywhere, and the unscrupulous barons of the food processing industry, will rejoice.

David Jones

The ins and outs of fertilization

SIR—Sperm gigantism in the genus *Drosophila* challenges our current understanding of the evolution of anisogamy¹. We have suggested previously that provisioning of the zygote by giant sperm may be a secondary adaptation, not a primary cause, of sperm gigantism². Here we demonstrate complex patterns of sperm utilization and sperm-egg interactions within the egg.

We examined sperm in fertilized eggs of 12 *Drosophila* species with three-dimensional reconstructions using anti-sperm antibodies and confocal microscopy. Sperm length was measured using a fully immersive, virtual-reality environment to trace sperm length^{3,4} (Fig. 1a, b, with the assistance of R. Brady). In *Drosophila ezoana* and *D. pachea*, eggs accommodate more than 15 mm of sperm, suggesting specific mechanisms for transport and compaction in the egg (Fig. 1c, d). Conversely, in two species with extraordinarily long sperm, *D. hydei* (23.2 mm) and *D. bifurca* (58 mm), only 1.31 and 1.6 mm of sperm, respectively, enters the egg (Fig. 1b, f), in contrast to the report by Bressac *et al.*⁵ that all, or most, of the sperm enters the egg in *D. littoralis*.

The portion of sperm not entering *D. hydei* eggs extended from the micropyle (the site of sperm entry) and lay on the exterior surface of the egg and substrate ($n=32$; observed by scanning electron microscopy, with the assistance of D. Johnson). This portion of the sperm in *D. hydei*, and perhaps all other species whose entire sperm does not enter the egg, is thus utilized by neither the egg nor the female.

The significance of the portion of sperm entering the egg is not known. It remains in the anterior region of the egg⁶ in all species, except the distantly related *D. busckii* (Fig. 1e), and displays species-unique, three-dimensional configurations. In most species, remnants of the axoneme persist within the midgut region of the late

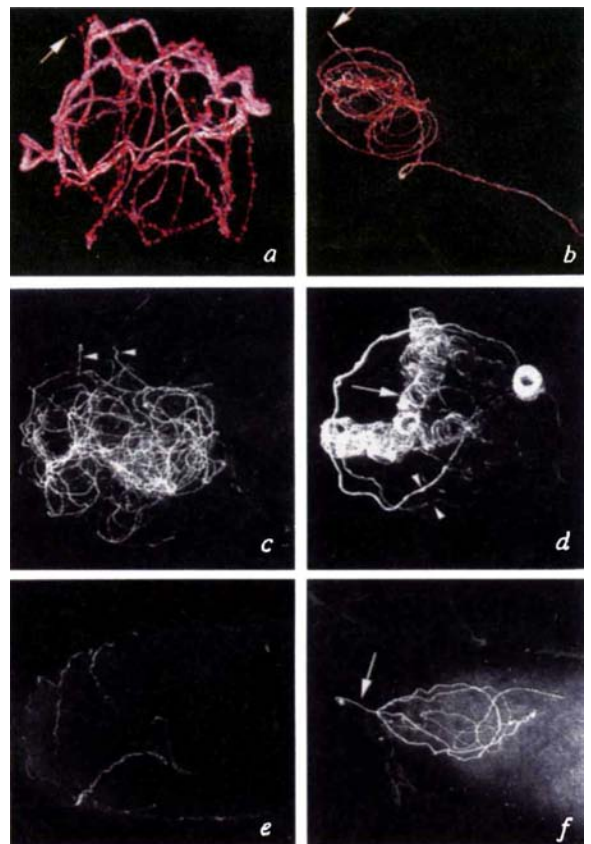
embryo (data not shown), although regional variations in staining intensity (Fig. 1c, d, arrowheads; ref. 6) suggest that some proteins are removed or diffuse from the sperm.

The genus *Drosophila* is divided into several subgenera⁷, within which giant sperm have evolved independently many times, primarily within the subgenus *Drosophila*⁸. The production of short sperm that completely enters the egg in *D. busckii* (subgenus *Dorsilopha*), *D. pseudoobscura*, *D. melanogaster* and *D. simulans* (subgenus *Sophophora*) suggests that this represents the ancestral condition (Fig. 2). The eight species from three lineages of the subgenus *Drosophila* that we examined display much variation in their pattern of fertilization, both among and within the lineages (Fig. 2).

In the first lineage, the entire sperm enters the eggs of *D. acanthoptera* and *D. pachea*, whereas only a fraction is incorporated in *D. nanoptera* and *D. wassermani*, a pattern independent of sperm length. The former species have sperm with an unusual helical morphology (Fig. 1d), which may represent an adaptation for compaction into the egg. In the second lineage (*D. littoralis* and *D. ezoana*), all or nearly all of the sperm tail was found in eggs; selection on males to provision zygotes through sperm may have contributed to sperm length evolution in this lineage. For the third lineage (*D. hydei* and *D. bifurca*), incorporation of only a fraction of the sperm into the egg is probably an ancestral trait; differences in sperm length (35 mm between *D. hydei* and *D. bifurca*) have thus evolved with no increase in the amount of sperm material entering the egg.

These data support theoretical models suggesting that males should not evolve a

FIG. 1 Visualization of sperm in fertilized *Drosophila* eggs by indirect immunofluorescence using an anti-sperm antisera essentially as described in ref. 6. Three-dimensional reconstructions were generated from volume-rendered confocal images (Zeiss LSM). Computer-assisted sperm length measurements in eggs fertilized by *D. wassermani* (a) and *D. hydei* (b) (two samples each from two independent experiments) were performed in the CAVE⁴, an immersive virtual environment, using CRUMBS (ref. 3). Tracings of sperm are shown in red connecting rectangular 'crumbs' placed on the image during the interactive session. A volume-rendered image of the confocal data set (light grey) is overlaid and lengths calculated by computer interpolation of the traced image data set. c, *D. ezoana*; d, *D. pachea*; e, *D. busckii*; and f, *D. bifurca*. Arrows indicate the site of sperm exit through the micropyle (removed during fixation). Arrowheads indicate examples of variations in staining intensity along the sperm. The antiserum used, AST-2, was one of a number prepared from mice injected with testes from *D. pachea*, *D. hydei* and *D. simulans*. All sperm were continuous from the male pronucleus to either the tip of the tail or the position of the micropyle. Sperm were measured from eggs fixed in a minimum of two independent experiments. At least 50 eggs from each species were scored for sperm tail structures, taken from collections containing >100 flies.



1. Parker, G. A., Baker, R. R. & Smith, V. G. F. *J. theor. Biol.* **36**, 529–553 (1972).
2. Pitnick, S., Spicer, G. S. & Markow, T. A. *Nature* **375**, 109 (1995).
3. Brady, R. *et al.* in *Proc. biomed. Visualization '95* (eds Loew, M. & Gershon, N.) 1825 (IEEE Computer Society, Los Alamitos, CA, 1995).
4. Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. & Hart, J. C. *Commun. Ass. Comput. Mach.* **35**, 64–72 (1992).
5. Bressac, C., Fleury, A. & Lachaise, D. *Proc. natn. Acad. Sci. U.S.A.* **91**, 10399–10402 (1994).
6. Karr, T. L. *Mechs Dev.* **34**, 101–112 (1991).
7. Grimaldi, D. A. *Bull. Am. Mus. nat. Hist.* **187**, 1–139 (1990).
8. Pitnick, S., Markow, T. A. & Spicer, G. S. *Proc. natn. Acad. Sci. U.S.A.* **92**, 10614–10618 (1995).
9. Parker, G. A. *J. theor. Biol.* **96**, 281–294 (1982).
10. Snook, R. R., Markow, T. A. & Karr, T. L. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11222–11226 (1994).

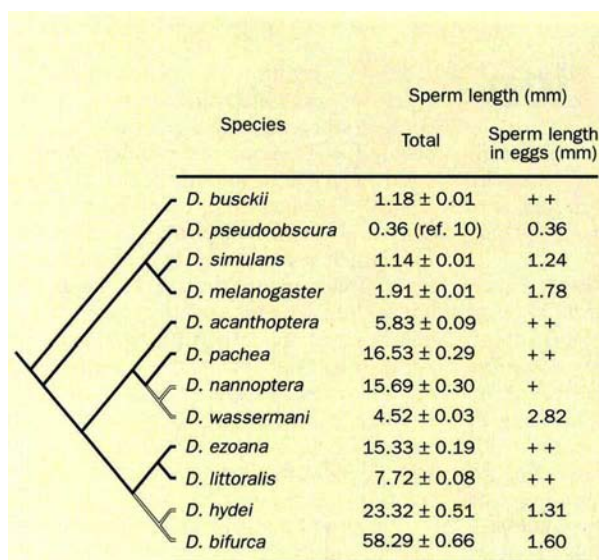


FIG. 2 Phylogeny, total sperm length and length of sperm entering the egg for 12 species of *Drosophila*. The phylogeny was reconstructed and total sperm length determined as described in ref. 8. Length of sperm entering the egg was either estimated or measured directly as in Fig. 1. In the six species for which complete tracing and measurement were not done, determination of whether nearly all (++) or only part (+) of the sperm enters the egg was based on visual observation of sperm in eggs and the position of the sperm tail relative to the egg micropyle. Solid branches indicate that entire sperm enters the egg; hollow branches indicate that only a fraction of sperm enters the egg.

provisioning function of their sperm when anisogamy ratios are large⁹, the ancestral character state for *Drosophila*.

Although our findings indicate that selection to provision offspring has not significantly contributed to sperm length evolution in *Drosophila*, they demonstrate a previously unrecognized diversity and complexity of sperm-egg interactions. An unappreciated degree of coordinated evolution between sperm and egg has occurred in *Drosophila*. Further studies

are needed to determine whether similar variation in sperm utilization occurs in other animal groups.

Timothy L. Karr

Department of Organismal Biology
and Anatomy,
The University of Chicago,
Chicago, Illinois 60637, USA

Scott Pitnick

Department of Biological Sciences,
Bowling Green State University,
Bowling Green, Ohio 43403, USA

Multituberculate phylogeny

SIR — In the recent reports of fossils of multituberculate mammals^{1,2} — pectoral girdle and ear ossicles, respectively — only a few selected characters and taxa are considered: 10 characters and 3 taxa in the report by Sereno and McKenna¹ (the remaining 12 characters are autapomorphies) and 16 characters and 4 taxa in that by Meng and Wyss². Basing phylogenetic hypotheses on a single complex of characters is sometimes necessary because of the fragmentary material available. However, the newly described material should be included in the more general framework of postcranial and cranial or basicranial comparisons (for example, refs 3, 4). We believe that additional taxa and character information are critical for the interpretations presented.

In their analysis, Sereno and McKenna¹ do not integrate the postcranial remains of the dryolestid *Henkelotherium*⁵ and the triconodont *Gobiconodon*⁶. They ally these taxa with multituberculates and therians, but give no rationale for the alliance. In fact, both taxa^{5,6} and other multituberculates (see, for example, refs 7, 8) show a higher degree of humeral torsion than do

therians and the multituberculate cf. *Bulganbaatar*¹. This conflicts with Sereno and McKenna's proposal of a single origin of parasagittal gait in multituberculates and therians. They also restore the scapulocoracoid of Early Jurassic *Morganucodon* without trace of a supraspinous fossa, yet this portion of the bone is not preserved⁹. Moreover, Tritylodontidae, an outgroup to *Morganucodon*^{3,4}, has a supraspinous fossa in the anterodorsal border of the scapula¹⁰, the same position at which the supraspinous muscle appears embryonically¹¹.

Meng and Wyss² offer four characters of the ear ossicles allying multituberculates with monotremes, but three of them are problematic. First, an incus articulating dorsally with the malleus, as occurs in monotremes, is yet to be documented in multituberculates. The only associated malleus and incus of a multituberculate (*Lambdopsalis*) are not dorsoventrally articulated¹². Second, a horizontal position of the ectotympanic is shared by monotremes and the multituberculate *Lambdopsalis*, but is also probably primitive for placentals¹³. Accepting a horizontal ectotympanic for multituberculates and

using the taxa in Meng and Wyss's tree, this character should diagnose Mammalia and not a Multituberculata–Monotremata clade. Nevertheless, indirect evidence suggests that a horizontal ectotympanic is not repeated in multituberculates other than *Lambdopsalis* and some Cretaceous taeniolabidoids with inflated vestibuli. The orientation of the oval window roughly parallels the ectotympanic in living mammals³ and thus predicts the horizontal ectotympanic of *Lambdopsalis*. Given this correspondence, an oblique ectotympanic is interpreted for most multituberculates, including paulchoffatiids and ptitodon-toids³. Third, the 'contact' between the ectotympanic and 'pterygoid' purported for *Lambdopsalis* is between two broken pieces, one from the ectotympanic, the other from the pterygoid. Moreover, in the platypus, the 'pterygoid' does not contact the ectotympanic, which is supported only by connective tissue¹⁴. Consequently, the character becomes equivocal, even accepting the tree topology proposed by Meng and Wyss². In addition, the free monotreme 'pterygoid' is a neomorph under any recent phylogenetic scheme; therefore, its homology with the multituberculate pterygoid is suspect.

The exercise of building a matrix with data combined from both reports yields one phylogenetic tree with the same topology as that presented by Sereno and McKenna¹, which is three steps shorter than the tree reported by Meng and Wyss². The anatomical information provided is welcome; however, evaluation of the competing phylogenetic hypotheses of multituberculate relationships must await the inclusion of additional taxa and characters.

Guillermo W. Rougier*

John R. Wible†

Michael J. Novacek*

*Department of Vertebrate Paleontology,
American Museum of Natural History,
New York, New York 10024, USA

†Department of Anatomical Sciences
and Neurobiology, School of Medicine,
University of Louisville,
Louisville, Kentucky 40292, USA

SERENO AND MCKENNA REPLY — Presley⁸ and Rougier *et al.* have commented on our recent description of a complete multituberculate shoulder girdle and our cladistic analysis which links multituberculates and therians on the basis of six synapomorphies in the shoulder region¹. As evidence against the therian-like structure and function of the multituberculate pectoral girdle and forelimb, they cite the greater degree of torsion in the shaft of the humerus of another multituberculate (*Lambdopsalis*¹⁵). Marked humeral torsion and fossorial habits, however, are clearly correlated among mammals (for example, moles among living therians). Increased humeral torsion in this avowed fossorial multituberculate from the

Palaeogene can thus not be interpreted with confidence as a "residuum of the primitive torsion between humeral head and elbow condyle"⁸.

Rougier *et al.* remark that we excluded other relevant fossil taxa from our analysis, such as *Gobiconodon*⁶ and *Henkelotherium*⁵. Regarding the former, our initial study¹⁶ showed that inclusion of its fragmentary pectoral girdle in the analysis has no effect on our results. Regarding the latter, which was published after our initial analysis, only a single, partial, crushed pectoral girdle is available. Finally, the small fossa on the distal end of the scapular blade in tritylodonts (absent in known multituberculates) may be homologous with some part of the broad crescentic fossa of therians, as Rougier *et al.* suggest. The presence of this distal fossa is, nevertheless, highly variable among therian outgroups (present in distant forms such as *Cynognathus*¹⁷, but absent in forms undeniably closer to mammals such as *Probainognathus* and *Probolesodon*¹⁸). In our comparative figure, we failed to note that the distal end of the scapular blade of *Morganucodon* is based on a complete scapula of a close relative, the tritheledontid *Pachygenelus*, which is remarkably similar to the former where these bones overlap.

We join Rougier *et al.* in encouraging future attempts to re-evaluate multituberculate relationships on a broader sampling of taxa and characters. Indeed, this is the only means available to test whether the structural changes we outlined in the pectoral girdle in multituberculates and therians actually constitute key evidence for their common ancestry.

Paul C. Sereno

Department of Organismal Biology
and Anatomy,
University of Chicago,
Chicago, Illinois 60637, USA

Malcolm C. McKenna

Department of Vertebrate Paleontology,
American Museum of Natural History,
New York, New York 10024, USA

MENG AND WYSS REPLY — Rougier *et al.*'s underlying objection stems from a misperception that our paper² and its comments concerning the placement of multituberculates was intended as an

exhaustive phylogenetic analysis of the major groups of mammals. The aim of our presentation of new anatomical information for the auditory apparatus in *Lambdopsalis*, in relation to several currently competing hypotheses of multituberculate relationships, was of necessarily limited scope. Attention was given to major taxa for which ear ossicles are currently known.

Moreover, their criticisms of 3 of the 16 anatomical features we presented are unsubstantiated and self-contradictory. First, the previously described fragmentary incus and malleus¹² are displaced (as noted in our Fig. 3 legend²). Thus, there is no basis for regarding the incus as having had a posterior position in *Lambdopsalis*. That the incus lies dorsal to the malleus (given the position of the fossa for the incudal articulation) is no more in doubt than the possession of a brain by mammoths — a supposition that has "yet to be documented" directly.

Second, Rougier *et al.* assert that the horizontal ectotympanic shared by *Lambdopsalis* and monotremes is primitive for placentals (and therefore Mammalia) as well. This contradicts not only a recent analysis arguing against the homology of this feature in therians¹⁹ and monotremes, but also their own polarity assessment of this feature³, in which an inclined ectotympanic was considered to characterize Theria ancestrally. Their assumption that orientation of the oval window predicts ectotympanic inclination has been contradicted repeatedly — not just by our specimen (V10777.3), in which the oval window inclines more than 30°, but also in other studies (for example, a ventromedially facing oval window in *Lambdopsalis*²⁰, a vertical ectotympanic coexisting with an inclined oval window in *Morganucodon*⁴, and a horizontal ectotympanic occurring with an anteroventrolateral oval window in *Scutisorex* (AMNH 48474)).

Third, the statement that the pterygoid is a broken piece (V10777.1) is incorrect. The widely appreciated problematic homology of the monotreme-multituberculate 'pterygoid' bone was signified in our paper by its placement in inverted commas in the figure legend. The claim that the monotreme 'pterygoid' is neo-

morphic is supported neither by phylogenetic studies^{3,4} nor by ontogenetic studies of the platypus skull^{21,22}. Even if one assumes this structure to be a neomorph in monotremes, this alone would not rule out its homology in multituberculates.

Finally, we would be the last to question the desirability of bringing to bear all available data in evaluating this, or any, phylogenetic question.

Jin Meng

Department of Vertebrate Paleontology,
American Museum of Natural History,
New York, New York 10024, USA

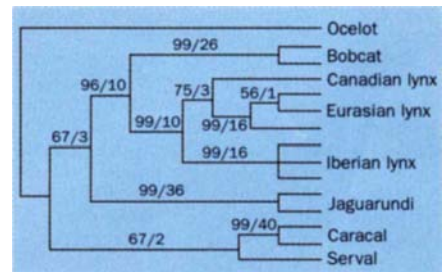
Andre R. Wyss

Department of Geological Sciences,
University of California,
Santa Barbara, California 93106, USA

Taxonomy of the Iberian lynx

SIR — The Iberian lynx, *Lynx pardinus*, is considered the most vulnerable cat in the world, yet its taxonomic status and relationship to other lynx species are controversial^{1,2}. Given that the Iberian lynx is listed as endangered and its populations are highly fragmented³, an understanding of its relationship to other taxa of *Lynx* is important for the development of an effective conservation plan. Here we report the first detailed molecular phylogenetic assessment of *Lynx* relationships. Our data suggest that the Iberian lynx is a distinct species relative to its European and North American counterparts.

The complete mitochondrial control region (D-loop) was sequenced for the Iberian lynx, Eurasian lynx (*Lynx lynx*), Canadian lynx (*Lynx canadensis*), bobcat (*Lynx rufus*) and related felid species (caracal, serval, jaguarundi and ocelot). Phylogenetic analyses of the D-loop performed using both maximum parsimony



Phylogeny derived using a maximum-parsimony analysis (branch and bound option) of aligned sequences. The single most parsimonious tree had a length of 410, a consistency index of 0.670 and a retention index of 0.791. Bootstrap values (1,000 replicates) are shown along branches, with the number of extra tree lengths needed to collapse a node separated by a slash. All specimens are unrelated individuals. The Iberian lynxes are from two different populations in Spain. Specific details of primers and experimental procedures are available from the authors on request.

- Sereno, P. C. & McKenna, M. C. *Nature* **377**, 144–147 (1995).
- Meng, J. & Wyss, A. R. *Nature* **377**, 141–144 (1995).
- Wible, J. R. *J. Vert. Paleont.* **11**, 1–28 (1991).
- Luo, Z. in *The Shadow of the Dinosaurs: Early Mesozoic Tetrapods* (eds Fraser, N. C. & Sues, H.-D.) 98–128 (Cambridge Univ. Press, New York, 1994).
- Krebs, B. *Berliner geowiss. Abh.* **133**, 1–110 (1991).
- Jenkins, F. A. Jr & Schaff, C. R. *J. Vert. Paleont.* **8**, 1–24 (1988).
- Kielan-Jaworowska, Z. & Gambaryan, P. P. *Fossils Strata* **36**, 1–92 (1994).
- Presley, R. *Nature* **377**, 104–105 (1995).
- Jenkins, F. A. Jr & Parrington, F. R. *Phil. Trans. R. Soc. Lond.* **B273**, 387–431 (1976).
- Sun, A. & Li, Y. *Verteb. palasiat.* **23**, 135–153 (1985).
- Cheng, C. J. *Morph.* **97**, 415–471 (1955).
- Miao, D. & Lillegren, J. A. *Natn. geogr. Res.* **2**, 500–507 (1986).

- Novacek, M. J. *Mamm. Rev.* **7**, 131–149 (1977).
- Gates, G. R., Saunders, J. C., Bock, G. R., Aitkin, L. M. & Elliot, M. A. *J. acoust. Soc. Am.* **56**, 152–156 (1974).
- Kielan-Jaworowska, Z. & Qi, T. *Verteb. palasiat.* **28**, 81–94 (1990).
- Sereno, P. C. & McKenna, M. C. *J. Vert. Paleont. Suppl.* **10**, 42A (1990).
- Seeley, H. G. *Phil. Trans. R. Soc. Lond.* **B186**, 59–148 (1895).
- Romer, A. S. *Breviora* **407**, 1–26 (1973).
- Zeller, U. in *Mammal Phylogeny Vol. 1* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 95–107 (Springer, New York, 1993).
- Miao, D.-S. *Univ. Wyoming Cont. Geol. Spec. Pap.* **4** (1988).
- Presley, R. & Steel, F. L. D. *Anat. Embryol.* **154**, 95–110 (1978).
- Zeller, U. *Die Entwicklung und Morphologie des Schädels von Ornithorhynchus Anatinus* (Kramer, Frankfurt, 1989).

and neighbour-joining resulted in identical topologies (see figure).

These results are significant for several reasons. First, the monophyly of the genus *Lynx* is strongly supported, a finding similar to that of Werdelin² but contrasting with some morphological accounts³. Second, relationships among the felid taxa are congruent with previous molecular data⁵. Third, the Iberian lynx is divergent from both the Eurasian lynx and Canadian species, and so can be considered an evolutionary unit (or valid phylogenetic species) in that from both a morphological and a genetic standpoint it is unique. The Iberian lynx revealed a lower level of nucleotide sequence divergence (0.006%) than the bobcat (0.61%) or Eurasian lynx (0.69%). It will be interesting to conduct a detailed study of the remaining isolated populations of Iberian

lynx in an effort to learn more about the overall phylogeographical pattern and levels of genetic variation in this rapidly vanishing species.

Juan F. Beltrán

*Estación Biológica de Doñana,
Consejo Superior de Investigaciones
Científicas, Apartado 1056,
41080 Sevilla, Spain*

John E. Rice

Rodney L. Honeycutt
*Department of Wildlife and Fisheries
Sciences,
Texas A&M University,
College Station, Texas 77843, USA*

1. Werdelin, L. *Ann. Zool. Fenn.* **18**, 37–71 (1981).
2. Corbet, G. B. & Hill, J. E. *A World List of Mammalian Species* (Br. Mus. Nat. Hist., London, 1986).
3. Rodríguez, A. & Delibes, M. *Biol. Conserv.* **61**, 189–196 (1992).
4. Salles, L. O. *Am. Mus. Novit.* **3047**, 1–67 (1992).
5. Janeczowski, D. N., Modi, W. S., Stephens, J. C. & O'Brien, S. J. *Molec. Biol. Evol.* **12**, 690–707 (1995).

Right-pawedness in toads

STR—Preferential limb use, of which human handedness is the most clear example, is a prominent aspect of brain lateralization and has been reported in both birds and mammals¹. There are reports of significant 'footedness' in avian species that use their feet to manipulate food and objects²; lateralized forelimb usage in large samples of inbred mice³; and in primates, recent

assessments have revealed handedness in several species⁴. Neural lateralization for control of vocalizations has recently been reported in anurans⁵, raising the possibility that preferences in limb use may exist in lower vertebrates, such as amphibians. Here we report the first evidence for behavioural asymmetry in forelimb usage at the population level in two species of

anurans. Our findings suggest that pawedness has a long evolutionary history, dating back at least to early tetrapods.

We have examined the forepaw used by European toads (*Bufo bufo*) of natural populations during attempts to remove either a plastic balloon wrapped around the head (expt 1) or a strip of paper stuck onto the mouth/nose region (expt 2). In both experiments, toads showed a bias for right forepaw use at the population level (see table).

It is often assumed that laterality of limb use evolved only in those species that use their limbs for manipulative activities⁶. Domestic chickens (*Gallus gallus*), however, do not use their feet to pick up or manipulate objects, but they use their right foot preferentially during ground scratching⁷. Thus, it has been suggested that it is not manipulative ability alone which confers a population bias of footedness in avian species, but rather active use of the feet for feeding or searching for food^{1,2}. *B. bufo* makes some use of the feet in feeding and also commonly wipes its head using one of its forepaws⁷. Active use of the forelimbs associated with feeding or grooming behaviour could thus produce asymmetrical forepaw usage in this species.

The South American cane toad (*Bufo marinus*) does not show asymmetry in the paper-strip test (see table; expt 3), but does have asymmetry in another behavioural test. We measured to which side *B. marinus* turned when positioned underwater, with the ventral surface of the toad uppermost (expt 4). From videotapes we determined that the toads pivoted preferentially to their left side. During the pivot, the left forepaw was released, and the right forepaw controlled rolling to the upright position.

Pawedness and motor asymmetries found in natural populations of toads could represent a precursor of handedness in higher vertebrates and thus contribute to our understanding of the evolution of brain lateralization.

Angelo Bisazza, Claudio Cantalupo

*Dipartimento di Psicologia Generale,
Università di Padova, 35131 Padova, Italy*

Andrew Robins, Lesley J. Rogers

*Department of Physiology,
University of New England,
Armidale, New South Wales 235, Australia*

Giorgio Vallortigara*

*Istituto di Filosofia, Pedagogia,
Didattica delle Lingue Moderne,
Università di Udine, 33100 Udine, Italy*

	Per cent 'right-pawed'	Student's t-test (two-tailed)	Probability	No. of animals showing preference		χ^2 test	Probability
				'Right-pawed'	'Left-pawed'		
<i>Bufo bufo</i>							
Expt 1 (n = 24)	59.2 ± 3.9	2.325	<0.05	14	4	5.55	<0.02
Expt 2 (n = 46)	55.2 ± 2.5	2.039	<0.05	26	10	7.11	<0.01
<i>Bufo marinus</i>							
Expt 3 (n = 18)	48.2 ± 5.5	-0.330	>0.1				
Expt 4 (n = 18)	66.4 ± 2.8	5.795	<0.001	15	2	9.940	0.002

Mature *B. bufo* were collected from natural populations (Valsanzibio, Colli Euganei, north Italy; expt 1: spring 1994; expt 2: spring 1995) and kept individually for at least 3 days before testing. Toads were placed in the middle of a circular tank (60 cm diameter) with a small plastic balloon wrapped around their head (expt 1), or a small wet piece of paper stuck on their mouth and nose (expt 2). They were given 10 successive trials; in each trial the first forepaw used in attempts to remove the annoyance was recorded. In expt 1 the animals were manipulated in turn by two experimenters (one right-handed and one left-handed); this produced no significant effect ($\chi^2(1) = 2.28$). In expt 2 the animals were manipulated in turn by two right-handed experimenters, different from those of expt 1 and unaware of its results. In both experiments experimenters were alternated, one placing the annoyance and the other keeping and placing the toad in the arena. Mean percentages of right forepaw use (with s.e.m.) and number of animals showing predominant right or left forepaw use are shown. *B. marinus* were collected from north Queensland (expts 3 and 4: autumn 1995) and group-housed for at least 1 week before testing. Paper-strip tests (expt 3), followed by submersion-inversion tests (expt 4), were performed on each toad for 6 successive days. Expt 3: toads were placed in a circular tank, with a small strip of wet paper placed across their mouth and nose equidistant from the midline. One trial per day was performed for each toad, the first paw used in attempts to remove the strip being recorded. Expt 4: each toad was turned upside down and, still clasping the experimenter's fingers, immersed in a small tank of water. The side to which the toads turned when righting was recorded. Three trials per toad per day were performed. Every effort was made to randomize the direction of rotation of the toad when inverting and the orientation of the toad with respect to the experimenter. On day 1, nine toads were tested with the left hand and the remaining with the right hand for each trial. On day 2, they were tested alternately with left or right hand (random sequence of toads). From day 3 onwards each toad was tested (in random order) with R, L, R or L, R, L hands, for each of 3 trials. Results show mean % of right forepaw use and left side turning (with s.e.m.), and number of animals showing prevalent left or right turning over the 6 days.

1. Bradshaw J. L. & Rogers L. J. *The Evolution of Lateral Asymmetries, Language, Tool Use, and Intellect* (Academic, San Diego, 1993).
2. Rogers, L. J. & Workman, L. *Anim. Behav.* **45**, 409–411 (1993).
3. Waters, N. S. & Denenberg, V. H. *Behav. Brain Res.* **63**, 195–204 (1994).
4. Ward, J. P. & Hopkins, W. D. (eds) *Primate Laterality: Current Behavioural Evidence of Primate Asymmetries* (Springer, New York, 1993).
5. Bauer, R. H. *Psychobiology* **21**, 243–248 (1993).
6. Walker, S. F. *Br. J. Psychol.* **71**, 329–367 (1980).
7. Ewert, J.-P. *Neuro-Ethology* (Springer, Berlin, 1980).

*To whom correspondence should be addressed.

A paean to living things

Christian de Duve

What is Life? By Lynn Margulis and Dorion Sagan. *Simon and Schuster*: 1995. Pp. 207. \$40.

A LITTLE over 50 years have gone by since the publication of the much acclaimed — and discussed — *What is Life?* by the theoretical physicist Erwin Schrödinger, a pioneer of wave mechanics who also kept an interest in the biology of his day (see the review by C. R. Welch in *Trends in Biochemical Sciences* 20, 45–48; 1995). Anyone who has read Schrödinger's sober and penetrating attempt to single out the key properties of life as he knew it might expect to find that a book of the same title published half a century later contained the carefully distilled quintessence of our present understanding of living processes, which is so much more profound and detailed than were the inklings available to Schrödinger. This, however, is not the goal the famous mother-and-son team have set themselves in this — their fourth — joint opus. There is hardly any mention of modern biochemistry, cell biology or molecular biology in this luxuriously produced extravaganza. We are treated, instead, to such definitions as: "Life is a material process, sifting and surfing over matter like a strange slow wave ... a controlled artistic chaos... a planetary exuberance... the watery membrane-bound encapsulation of spacetime... the strange new fruit of individuals evolved by symbiosis... a network of crosskingdom alliances ... the transmutation of sunlight."

Remarks of this sort conclude each of the nine chapters in which the authors, after a number of introductory generalities, sweep over the four-billion-year history of life on Earth in a smooth, vivid style, which, however, often eschews factual complexities in favour of anecdotal information, poetic imagery or philosophical digressions. Readers acquainted with the authors' previous works will recognize

some familiar themes. There is the impassioned defence of the much-maligned bacteria, which the authors see as identified in the public mind only with the production of diseases, the fervent allegiance to James Lovelock's Gaia hypothesis and, especially, Margulis's main contribution to science, the endosymbiotic theory of the origin of

ENTOMBED in stone — Carboniferous limestone containing the fossilized remains of crinoids. These primitive echinoderms have delicate feathery arms radiating from a central stem made of calcite plates, or ossicles, that, as shown here, often break off from the fossils. From *An Illustrated Guide to Fossils* by Chris Pellant. Clearly written, well organized and abundantly illustrated. Dragon's World, £19.95.



eukaryotic cells. This she presents in her customary truncated version which leaves out such important eukaryotic features as the cytomembrane system, cytoskeletal elements other than microtubules (which she believes to have been brought in by symbiotic spirochaetes) and, especially, phagocytosis, notwithstanding the probable role of this process in endosymbiont adoption. These themes are entertainingly woven into a patchy, almost pointillist narrative, enlivened by sumptuous pictures and colourful vignettes drawn from the authors' rich lore of biological oddities. Those who know the gist of the topic will find many choice bits to enjoy. How much uninformed readers will discern of the realities behind the metaphors is not clear.

They will get the philosophical message, however, and it is one that most scientists will receive with some surprise, if not dismay. Purposefulness, consciousness and freedom to choose are identified as properties common to all living beings. In a discussion of chemotaxis, for example, microbes are said to "make selections", to "choose". An amoeba, deprived of its favourite food, turns "reluctantly" to less delectable fare. This is not just poetic

licence. "At even the most primordial level," we are told, "living seems to entail sensation, choosing, mind.... All [organic beings] are sentient, possessing the internal teleology of the autopoietic imperative Life is matter that chooses."

In the authors' depiction of evolution, Darwinian selection plays second fiddle to sentient, willful, goal-oriented selection. Darwin himself, in their lengthy discussion of his dispute with Samuel Butler, comes out as something of a fraud, guilty of ignoring the contributions of others, including his own grandfather Erasmus, for the sake of self-promotion. Margulis and Sagan clearly bend in favour of Butler rather than Darwin, to the point of expressing sympha-

thy with Butler's Lamarckian view that the consciously learned may eventually enter the realm of the hereditary unconscious, although they concede that "it seems doubtful that children will ever be born reading and writing".

In an epilogue to his *What is Life?*, entitled "On Determinism and Free Will", Schrödinger rejected Western philosophies in favour of Vedantic beliefs. The Chilean biologist Francisco Varela, who is warmly commended by Margulis and Sagan for his "autopoietic" view of life as an "emergent holarchy", has embraced Buddhism. Although Margulis and Sagan do not advocate any particular creed, their brand of anti-Darwinian 'bacterio-mysticism' will no doubt be hailed by many a creationist, as well as by all those who, for one reason or another, are disenchanted with contemporary science, which they accuse of being overly materialistic, deterministic and reductionist. □

Christian de Duve, emeritus professor at the University of Louvain, Belgium and at the Rockefeller University, New York, is at the International Institute of Cellular and Molecular Pathology, 75 Avenue Hippocrate, B-1200 Brussels, Belgium.

■ Christian de Duve's airs some further thoughts on the origin of life in the recently published *What is Life? The Next Fifty Years* edited by Michael P. Murphy and Luke A. J. O'Neill. The book arose from a conference held in 1993 to celebrate the fiftieth anniversary of Schrödinger's original book of the same title. Other contributors speculating on the future of life include Jared Diamond, Manfred Eigen, Stephen Jay Gould, John Maynard Smith, Roger Penrose and Stuart Kauffman. Cambridge University Press, £17.95, \$24.95.

Social complexities

W. C. McGrew

In Quest of the Sacred Baboon: A Scientist's Journey. By Hans Kummer (translated by M. Ann Biederman-Thorson). Princeton University Press: 1995. Pp. 337. \$29.95.

In the field study of monkeys and apes, a species often becomes linked with a particular scientist, such as chimpanzees with Jane Goodall or orangutans with Biruté Galdikas. This association sometimes reflects a life's work of journal articles, conference presentations and book chapters synthesized into a single academic volume, plus a memoir of the same

takes the reader straight into the bush, to the moving final paragraph that will resonate with long-term field workers, it is compelling.

Hamadryas baboons are worthy of attention on several fronts. They are probably the toughest of all higher primates, surviving in the hottest, driest and most open spaces, where sparse resources such as water and shade can be a matter of life or death. Their society has four hierarchical layers: the troop, numbering hundreds, reflects the scarcest resource of all, the cliffs for safe overnight sleeping; the band, numbering scores, is suited to other defensible resources such as a water-hole; the clan, numbering tens, relates to the richest food-patches such as 'fields' of underground tubers; finally, the family of typically a fully adult male, several adult

Bruce Coleman

females and their offspring, plus a secondary, 'follower' male, is the smallest unit, linked to the patchiest resource, such as the shade or fruits of a single tree. Kummer argues that such complex social life is an adaptation to these most testing of environmental conditions; the simpler, troop-oriented lives of the rest of Africa's savanna or forest baboons simply would not suffice. That such social structure is inherited (and

this persists in zoos, even when captive baboons are emancipated from these constraints) became clear when Kummer's team studied natural hybrids: where the range of hamadryas and savanna baboons overlap, their societies show some intermediate characteristics.

Kummer's view of baboonery is arguably unique. Unusually in primatology, he is a classical ethologist, and so interested not just in the function of behaviour, but also in its causation, development and evolution. Some of his keenest insights apply to how (rather than why) individuals succeed in such a stratified society. Also, he has been committed from the start to combining many different methods: observation and experimentation in captivity and nature. It is the latter combination that is most impressive: replicating controlled encounters between triads of such dentally armed, large mammals in the middle of the bush is no picnic. Of course there is also fuel for debate: he rails against anecdotes, but recounts some excellent ones; he warns against anthropomorphism, yet characterizes the species' core relationship as marriage; he generalizes about primates, when he really means Old World monkeys and apes.

The book has apt figures, including eight colour plates. Citation is scanty, and key figures such as John Crook are mentioned but not referenced; more puzzling, the

index is curious in that co-workers such as Angst and Kurt are absent whereas Sigg and Stolba are present. There are two reference lists, one being the complete bibliography of scientific publications on Project Hamadryas by the Zurich group, should the reader wish to pursue details; the other list more conventionally refers to all other works cited.

The pertinent question for the reader of *Nature* is whether or not the book is too specialized for the non-primatologist. The answer is that the scientific material is as richly focused as one would expect, but the other half of the book, the thoughtful, personal disclosures, are uncommonly telling, and deserve wide attention. □

W. C. McGrew is in the Departments of Sociology, Gerontology and Anthropology, and the Department of Zoology, Miami University, Oxford, Ohio 45056, USA.

Bending perception

Irving M. Klotz

Nazi Science: Myth, Truth, and the German Atomic Bomb. By Mark Walker. Plenum: 1995. Pp. 325. \$28.95.

THE jacket of Walker's new volume has the first two words of its title printed in inch-high letters, the next four in quarter-inch letters and the final three in half-inch letters. But, there really was no "Nazi science", except perhaps for some pseudo-scientific ramblings such as Hanns Hörbiger's cosmic ice fantasies, which impressed Himmler and Hitler. What did exist were scientists who were Nazis. Despite its grandiloquent title, this assembly of previously published material is essentially a collage of three topics (dealing almost solely with German physicists) related primarily by historical contiguity. Scintillating slogans ("there are neither simple answers nor simple questions") are interspersed as guides for the perplexed. This monograph would have benefited substantially from greater care in preparation, in minor matters as well as major ones. An example of a trivial item is the statement that the president of Germany in 1934 was Otto von Hindenburg, when his actual name was Paul Ludwig Hans Anton von Beneckendorff und von Hindenburg. Far more serious faults are the misleading and erroneous interpretations of many aspects of fission physics and of the state of understanding of that field by the wartime German atomic scientists.

About a quarter of the book is devoted to a chronicle of the machinations and rantings of the Nobel prize-winner Johannes Stark, who had joined the Nazis in the early 1920s. He was a frustrated, bitter man with an exceptional talent for

IMAGE
UNIMAGEABLE
FOR A COPYRIGHT
FOR A COPYRIGHT
REASONS

Toughing it out — is the complex social life of the hamadryas baboon an adaptation to harsh environmental conditions?

material for a more popular audience. Here, the title hints at an innovative attempt to combine the two, and Hans Kummer delivers in superb fashion.

His species is *Papio hamadryas*, the sacred (to the ancient Egyptians) or desert baboon, as studied mainly in the arid scrublands of eastern Ethiopia. Originally published in 1992, the book's German title, *Weisse Affen am roten Meer: Das soziale Leben der Wustenspaviane*, adds more: another population of these monkeys lives across the Red Sea in Saudi Arabia. (Kummer may be the only primatologist to study wild subjects of the same genus on two continents.) His 'Project Hamadryas' spanned 20 years, starting in Zurich Zoo and finishing in 1977 when hostilities in Somalia made the region too insecure.

The resulting volume is truly multi-dimensional. On the one hand it is a distillation of decades of making sense of the most complicated social organization of any species of non-human primate. This is achieved with elegance and economy, culminating in a six-page summary chapter that is a paragon. On the other hand it is the reflections of a 65-year-old Swiss scholar, ranging from revelations of conscience (for example, he regrets some of his early experiments, on ethical grounds) to aspects of his non-scientific life that vary from litterbug to gunslinger. From the arresting first page of the preface, which

irritating and antagonizing almost everyone, including Nazi officials. Walker's description confirms the general view that Stark should be allowed to sink even deeper into the oblivion he has so richly earned.

Another quarter of this monograph contains an extensive review of the increasing pressures from the Nazi bureaucracy and its supporters in the scientific community to mould institutions, such as the Prussian Academy of Sciences, into instruments to further their goals. According to Walker, various despicable actions by the Prussian Academy effectively destroyed it as a scientific enterprise. With this background, he embarks on some journalistic moralisms, based on his banal insight that, on a continuum from brutish criminal to virtuous saint, it is difficult to draw a sharp line differentiating Nazi from anti-Nazi. Nevertheless, he proceeds to formulate judgements on various individuals "according to objective criteria", and opines that even a heroic figure such as Max von Laue did not come through the Nazi period without a blemish, having bent slightly to increasing Nazi pressure. One might note that the rare individuals such as Carl Ossietzky and Dietrich Bonhoeffer who remained unbent were "*vernichtet*" (annihilated) by the regime.

The largest portion of Walker's book focuses on the German atomic bomb effort. In rhetorical space, words often have a fluidity that allows them to be fitted into a container of any shape. According to Walker, Ernst von Weizsacker and Werner Heisenberg "knew" of element 94 (plutonium) before 1945, for it is mentioned in passing in secret wartime documents from the German Uranium Club. There is even more convincing evidence that most American graduate students in nuclear physics or nuclear chemistry at that time "knew" about plutonium, for a contemporary textbook, written in 1940–41 and printed in 1942, states that "it was proved by McMillan and Abelson that the resulting 93^{239} [from U^{239}] decays to 94^{239} ". But how do these 'knews' compare with the discussion in Louis Turner's manuscript, submitted in May 1940 to *Physical Review* (and printed in 1946), containing a detailed analysis of how to convert "all of the U^{238} into Pu^{239} "? Clearly the word 'knew' has widely disparate meanings that need elaboration before one makes startling claims.

Of greater gravity, however, are some of the styles of argument used by the author. Walker employs a mode that has been trenchantly described by the expres-

sion (created by the classicist Mary Lefkowitz) *si potest esse, est*: if something can be imagined to have happened, then it did indeed occur. Walker is anxious to convince people that Heisenberg really understood how to calculate the correct critical mass of an atomic bomb. To support this thesis, he cites a secret 1942 report in which this crucial number is stated to be between 10 and 100 kg, which is near the correct answer. The report gives no indication of the source of this

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Heisenberg's uncertainty: admitted to never having worked out the critical mass of an atomic bomb.

estimate or how it was reached. Nevertheless, we are told, "it seems most likely that Heisenberg would have been entrusted with this task". Consequently, we should conclude that indeed he must have carried it out.

One might think that it would be relevant to mention that there are other German secret wartime documents, some originating from Heisenberg, that consistently give the critical-mass values as tons. Furthermore, the main section of the 1942 report cited by Walker also gives a figure of tons. What is even more damaging to Walker's version is that in the Farm Hall transcripts, surreptitiously recorded conversations (declassified in 1992) of the German atomic scientists detained in England from May to December 1945, the critical masses given by Heisenberg and his colleagues encompass the following values: 2 or 4 or 10 or 50 or 180 or 200 or 400 or 500 or 5,000 kg, or 1 or 2 or 10 tons. In addition, at one point Heisenberg ultimately says: "Quite honestly I have never worked it out." One might think

this statement rules out Walker's interpretation of the secret 1942 report. But no, he simply adds another *si potest esse, est* on to the first one: Heisenberg must have *forgotten* that he had previously made the calculation.

In support of the latter conclusion, Walker cites the theoretical calculation of the critical mass recorded in the Farm Hall transcript of the lecture Heisenberg gave to his fellow detainees a week after the announcement of the Hiroshima explosion. At that point Heisenberg *knew* that a bomb could be constructed. Furthermore, Allied announcements essentially disclosed the size of the bomb. In Walker's judgment, Heisenberg's lecture is a "surprisingly accurate speculation" on the physics of the bomb. However, when that lecture is compared with the discussion of the bomb theory by Robert Serber in the *Los Alamos Primer* in March 1943, a diametrically opposite conclusion is reached (Serber and Logan): "[the Heisenberg] lecture is a primitive, unsophisticated treatment of the subject, treating a bomb as though it were a reactor."

In his rendition of the contents of the Farm Hall transcripts, Walker uses another stratagem to support his judgements. He presents his edited versions of the statements instead of direct quotations. This approach can give misleading impressions. For example, Walker says that after learning of the detonation of the bomb, "The detainees... at first... did not believe their English wardens." What actually appears in the transcripts after their 'host' Major Rittner told the 'guests' that a bomb had been exploded is the following:

Heisenberg: "It's got nothing to do with atoms... some dilettante in America... has bluffed them into saying 'if you drop this it has the equivalent of 20,000 tons of high explosive' and in reality doesn't work at all.... I don't believe a word of the whole thing... I don't believe it has anything to do with uranium."

These words show that Heisenberg was utterly incredulous, not just surprised, and lead to a very different view of what Heisenberg really knew.

Another example of how misleading edited selections from the Farm Hall transcript can be is provided by Walker's statement that "the Germans had been discussing elements 93 and 94... throughout their captivity". This is literally true but misleading, for it does not point out that the main focus and largest amount of attention was devoted to element 91 as the likely fissionable explosive in the Allied "plutonium" bomb. The essence of that discussion appears in Heisenberg's assessment:

"I believe that it is almost more likely that they have done something quite original such as getting out protactinium from colossal quantities of material.... Perhaps

the facts are that they thereby discovered 'Pluto'. Pluto is a code name. Protactinium also starts with a 'P'.... Perhaps the others have used protactinium: this is almost easier to imagine than all other methods.

What a different light this quotation throws on what Heisenberg and his colleagues "knew" about element 94.

Churchill is reputed to have said "man will occasionally stumble over the truth, but usually manages to pick himself up, walk over or around it, and carry on". How apt a description of Walker's writings on the German atomic bomb project. ▲

Irving Klotz is in the Department of Chemistry, Northwestern University, Evanston, Illinois 60208-3113, USA.

Who enforces the law?

Giovanni F. Bignami

Idee per il governo: L'Università. By Raffaele Simone et al. Laterza: 1995. Pp. 187. L15,000.

Il peso della qualità accademica. By S. Garbisa and L. Calza. Cleup: 1995. Pp. 123. L18,000.

WE are fortunate to have Laterza's series of books on 'Idee per il governo'. The latest addition, a small paperback on Italian universities, is a natural follow-up to Carlo Bernardini's earlier volume on scientific research in Italy. The author of the present tome is Raffaele Simone, an old hand in castigating Italian academic institutes and their members. The book could hardly be more timely (no doubt by chance): for the first time, about 50 *concorsi* (nationwide competitions for professorships) are being scrutinized by magistrates for alleged corruption and abuse of office — a scandal recently given wide coverage by the Italian mass media. This is a welcome development: until now, spectacular unfairness and even gross fraud have been consistently perpetrated with guaranteed impunity.

Nature has championed a just cause by publishing reports of several of these scandals, as well as correspondence on the subject, notably from Fernando Aiuti and colleagues and from Spiridione Garbisa and Laura Calza (the authors of the second book under consideration). Two recent news items are particularly enlightening, particularly on the end of what used to be a true *omertà* wall (see *Nature* 374, 756 & 378, 228; 1995).

Simone does more than simply stimulate an analysis of what went wrong — and is still wrong — with Italian universities; in convening a group of experts, in the poignant style of the series, he also brings

forward proposals for improvement. A possible starting point is suggested in the contribution by Tullio De Mauro, who makes a simple but arresting observation. There are currently 85 institutions across the world that have survived more-or-less unchanged for over five centuries. Among these are the Roman Catholic Church, the British parliament, several Swiss cantons and the Monte dei Paschi of Siena. But most (more than 70) are universities, many in Italy.

Even in Italy, the message is that universities are here to stay, although sooner or later they will have to be brought, for good or bad, in line with the evolution of Italian society. To this end, Simone lists 11 detailed tasks facing the reformer in raising the Italian university to the level he sadly calls "just about decent". It seems getting there will be a long uphill trek, given the lowlands we are in now.

Another contributor is Antonio Ruberti, a former minister for universities and European commissioner. During his short ministry, Ruberti succeeded in having approved by parliament a package of four laws providing a framework, and some political guidance, for raising the Italian university system to a modern, European level. Some five years later, they are still largely waiting to be applied. It would seem that the cause is the usual inextricable battle for responsibility between the ministry in Rome, always fearful of losing control (that is, power), and individual universities, autonomous sometimes in thought but rarely in practice. Nowhere is Dante's splendid hendecasyllable "Le leggi son, una chi pon mano ad esse?" (poorly rendered, "Laws do exist, but who'll stand to enforce them?") more applicable.

Or is the culprit simply indifference, in a political and media environment that has lost all sense of perspective for anything cultural? Riccardo Chiaberge, a brilliant columnist of the *Corriere della Sera*, raises this suggestion. The words *autonomia universitaria*, he argues, are by now a kind of "mantra", an "*om ma ni pad me hum*" (erroneously here interpreted as sanskrit), mechanically repeated without pause to grasp its actual meaning.

All the contributors agree that too much visibility is given to the *concorsi* scandals, and that this exposure artfully hides deeper, more fundamental problems in the higher-education system of a country still coming up for air after 20 years of fascism followed by 40 years of uninterrupted, rigorous rule by the Christian Democrats.

Nevertheless, professors have to be enrolled, and the fairer the means, the better. Here Garbisa and Calza's book should come in handy. It is, as the title says, a tool for judging academic quality through the use of "objective parameters". This little volume should be read literally to the last page: the last four contain an illuminating and complete bibliography on the *concorsi*

topic as well as on justice (or injustice) in Italian universities, listing a total of 76 items. (Here *Nature*, with 23 entries, scores about as high as all the Italian daily and weekly publications together.) Most Italian scientists are familiar with such parameters as the citation index and impact factor. But with alarming frequency these notions seem to escape their minds when it is their turn to sit as members of a *concorso* commission. So it is appropriate, then, that the authors remind everyone that there are methods for an objective evaluation of candidates by competent and honest peer review, that no magic is needed and that commissions, if honest, simply cannot commit gross injustice if they follow such methods. They add a selection of clear examples to avoid misunderstandings by those who still find such modern intricacies difficult to grasp.

Gilberto Muraro, *rector magnificus* of Padua University (one of the venerable institutions mentioned above) makes it clear: that it is only on the basis of bibliometric methods such as those described in this booklet that we can ever hope to enrol professors fairly and, in the end, fruitfully. No desperate, last-ditch battle for every *concorso*, but also no room for the dreaded *ope legis*, promotion by virtue of law, always lurking in some of the dark corridors of power. □

Giovanni F. Bignami is at the Istituto di Fisica Cosmica CNR, 15 via Bassini, 20133 Milano, Italy.

Correction

The correct spelling of the name of the editor of *Aristophanes' Birds* reviewed by Malcolm Davies in *Nature* 379, 307–308; 1995) is Nan Dunbar.

New in paperback

Facing the Future: The Case for Science by Michael Allaby. Bloomsbury, £7.99. A passionate effective counterblast to the recent public backlash against science. The author is especially good on environmentalism, his home turf, exposing as nonsense many current fears about global warming, nuclear power, pesticides and pollution.

The Spirit of the System: Lamarck and Evolutionary Biology by Richard W. Burkhardt, with a new opening by the author. Harvard University Press, \$17.95, £11.50. "It is a great relief to have at last a good book on Lamarck in English, neither partisan through patriotism, nor distorted by dislike of Darwin. Burkhardt gives us an excellent introduction to Lamarck, his contemporaries and predecessors, and to the scientific bodies and institutions in revolutionary Paris... Exceptionally well written" wrote A. J. Cain in a review of the original edition in *Nature*.

Electrons in artificial atoms

R. C. Ashoori

Progress in semiconductor technology has enabled the fabrication of structures so small that they can contain just one mobile electron. By varying controllably the number of electrons in these 'artificial atoms' and measuring the energy required to add successive electrons, one can conduct atomic physics experiments in a regime that is inaccessible to experiments on real atoms.

THE puzzle of atomic spectra was a prime motivation for the development of quantum mechanics. Niels Bohr unravelled the mystery by determining that the wavelike nature of electrons allowed them to occupy only discrete orbits within an atom, with well defined energies. Starting about 10 years ago, advances in semiconductor technology allowed the fabrication of structures so small that their discrete quantum level structure was resolvable. In the past few years, powerful new spectroscopic probes have revealed a wealth of new physics in these 'artificial atoms'.

Essentially, artificial atoms are small boxes about 100 nm on a side, contained in a semiconductor, and holding a number of electrons that may be varied at will. As in real atoms, electrons are attracted to a central location. In a natural atom, this central location is a positively charged nucleus; in an artificial atom, electrons are typically trapped in a bowl-like parabolic potential well in which electrons tend to fall in towards the bottom of the bowl. One can consider the artificial atom as a tiny laboratory in which quantum mechanics and the effects of electron-electron interactions can be studied.

The spacing between atoms in a semiconductor crystal is typically about 0.3–0.4 nm. In artificial atoms, electrons are confined in structures about 100 nm in diameter. Thus, an artificial atom in a crystal comprises many real atoms. The quantum-mechanical theory of solids explains why the electrons do not get trapped on the real atoms of the crystal and instead only sense the potential well of the artificial atom. This theory dictates that some electrons in a crystal behave as free electrons, albeit with a different mass¹. For example, electrons in the semiconductor gallium arsenide appear to carry a mass that is only 7% of the mass of free electrons.

The physical characteristics of an artificial atom (sometimes called a 'quantum dot') differ considerably from that of a natural atom for one important reason: artificial atoms are typically much larger than real atoms. The electron orbits do not simply scale with size. Imagine an atom containing many electrons whose size is continuously variable. As it is made larger, the Coulomb energy arising from the repulsion between electrons orbiting around the nucleus decreases because the average spacing between electrons increases. However, there is another energy scale in the problem: the separation in energy of the different orbits of electrons in the artificial atom. As the atomic size increases, the differences in the orbital energies decrease faster than the Coulomb energy. It follows that in a large atom, the effects of electron-electron interactions are relatively more important than in a small atom.

It may not be long before these distinctions affect modern electronics. The present view of small electronic devices depicts them as controlling the motion of small and classical 'seas' of a few thousand electrons. As devices shrink, this view no longer holds. In fact, it is already possible to create electronic devices small enough to have device characteristics sensitive to the motions of single electrons within them, even at room temperature^{2,3}. Electronic devices may no longer be seen as small seas of electrons but instead as very large 'atoms'. Building devices at this size scale that

have reproducible and desired electronic properties will be an immense challenge. This Review describes structures in which researchers have now succeeded in precisely controlling the number of contained electrons, down to as few as one electron. It focuses on the unusual physical features of these large atoms unveiled by these studies. At low temperatures, electrons fall into distinct quantum-mechanical energy levels of these atoms, and the large Coulomb energy exerts a profound influence. Transitions never observed in the spectra of 'natural' atoms are readily seen for artificial ones. The electrons in artificial atoms can also act lethargically, reluctant to displace themselves to make way for another electron, even on a timescale of milliseconds. Most strangely, the large Coulomb repulsion can make it seem that electrons attract one another.

Measurements of the orbital energies of electrons in an artificial atom thus yield a wealth of data about a new physical regime. A large number of experimental probes have been used. Some techniques are electrical, such as measuring the current through a single dot^{4–10}, or the capacitance of arrays^{11–13} and single quantum dots¹⁴. Others have employed infrared and optical spectroscopies on arrays of dots^{15–17} and on individual dots¹⁸.

Construction and measurement

I will focus here on results from two newly developed techniques^{5,14,19} which have particularly affected our understanding of quantum dots. These methods are unique in allowing extraordinarily high-resolution spectroscopy of single quantum dots. The energy resolution of these spectroscopies is limited solely by the sample temperature. Both types of experiments have been performed on artificial atoms constructed inside a gallium arsenide semiconductor and at very low temperatures (0.05–0.30 K).

Figure 1a shows schematically the type of sample used in one of these experiments. The quantum dot is located between two capacitor plates. It is close enough to one of the plates to allow single electrons to tunnel (or hop) between the artificial atom and the nearby plate. Tunnelling is a quantum-mechanical process that allows electrons to pass through barriers that would be impenetrable classically. The artificial atom is far enough from the other capacitor plate to prohibit tunnelling to this plate.

How is such a sample actually realized? Utilizing modern semiconductor technology, it is possible to grow gallium arsenide (GaAs) semiconductor crystals one atomic layer at a time²⁰. Moreover, the material composition can be changed in successive layers. Some of the gallium in a layer can be replaced with aluminum to create aluminum gallium arsenide (AlGaAs). AlGaAs acts as an insulator, whereas electrons move freely in GaAs. By sandwiching a 10-nm thickness of GaAs between AlGaAs layers, one confines electrons in a GaAs 'quantum well'. This quantum well is so thin that at low temperatures, only the lowest quantum energy state of the well is occupied by electrons. The electrons have no freedom to move in the direction perpendicular to the well, but may only move laterally in it. By

placing electrostatic gates on the surface of the wafer, we can laterally confine this two-dimensional electron gas and create a quantum dot.

Figure 1b demonstrates how the schematic picture shown in Fig. 1a is realized in an experiment. A crystal wafer is grown layer by layer, starting from the bottom of the diagram. The first layer is a silicon-doped GaAs layer. The silicon doping makes this layer metallic, and it acts as the bottom capacitor plate of Fig. 1a. Then a thin (10-nm) AlGaAs barrier layer is grown. This barrier is thin enough that electrons can leak through it. Above that, there is a GaAs quantum well, and grown on top of the well is a thick (not leaky) AlGaAs barrier. On top of the crystal wafer, chromium is deposited to form the top capacitor plate, which I will refer to as the 'gate'. Additional sample processing^{21,22} on the sample surface is used to create a gate which laterally confines electrons in the quantum well below, creating a quantum dot.

Electric fields can be created by applying a voltage between the plates of the capacitor. If the top plate (gate) is made positive compared to the bottom one, electrons from the bottom plate will be attracted in the direction of the top plate, towards the artificial atom. Single electrons can thus be coaxed to tunnel into the dot or expelled from it through the application of voltages on the top plate.

After each electron is added to the dot, additional gate voltage is usually needed to coax another electron onto the dot. As I discuss below, this happens partly because after each electron is added to the dot, the electron charge in the dot is larger, and other electrons are repelled from entering. There are therefore specific gate voltages at which an electron is added to the dot, and it is this spectrum of voltages (the electron addition spectrum) which has been particularly useful in understanding quantum dots.

The motion of the single electrons into or out of the dot can be detected using a simple physical principle. When a single electron tunnels into the artificial atom, it moves closer to the top plate of the capacitor, and electrons in the top plate tend to be pushed away from the plate; that is, some charge is induced on the top plate. In these samples, the amount of charge induced is about half of an electron's charge. A specialized transistor enables detection of this charge and thereby allows measurement of the gate voltages at which single electrons were added to the artificial atom. A small a.c. voltage of frequency around 100 kHz is added to the d.c. gate voltage. When the gate voltage is adjusted to a voltage at which an electron can be added to the dot, the a.c. voltage causes the electron to tunnel back and forth between the dot and the bottom plate. At these gate voltages, charge appears synchronously with the a.c. voltage. A synchronous detector then registers a signal only at these voltages, yielding a peak in the response. The scheme measures the capacitance signal due to a single electron and is known as single-electron capacitance spectroscopy (SECS).

Results from this method, taken on a sample cooled to 0.3 K, are shown in Fig. 1c. The first peak on the left of the graph corresponds to the first electron in the quantum dot. Subsequent peaks correspond to additional electrons added to the quantum dot. The peak structure is completely reproducible as one scans the gate voltage up and down, and the widths of the peaks directly reflect the temperature of the sample.

The other type of experiment is performed on the type of sample displayed in Fig. 2a. Like the SECS method, this experiment also measures the gate voltages at which electrons are added to the dot. The detection method is however, completely different. Two current leads are in such close proximity to the dot that electrons may tunnel between the leads and the dot. A small voltage difference is applied between the left and right leads, and the experiment consists of monitoring the current through the quantum dot as the gate voltage is varied. I will refer to this type of experiment known as 'gated transport spectroscopy' (GTS).

For a current to flow, single electrons must pass through the quantum dot. This can only happen under special conditions. Just as in the SECS experiments, at particular gate voltages, it becomes energetically favourable for one more electron to be added to the

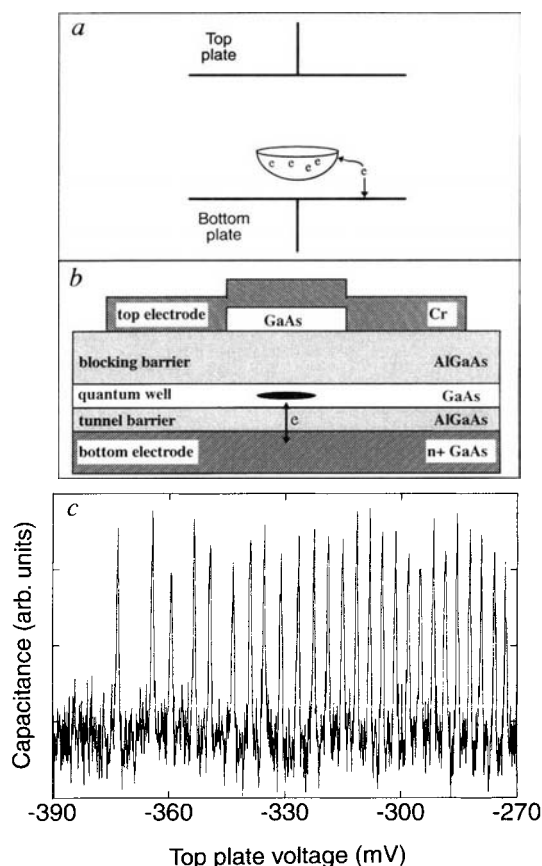


FIG. 1a, Schematic diagram of an artificial atom located between two capacitor plates. The artificial atom is actually two-dimensional; the bowl-like shape is to represent the force tending to move electrons to the centre of the atom. b, Diagram of the sample used in single-electron capacitance spectroscopy (SECS) experiments in a crystal grown using molecular-beam epitaxy. The artificial atom is the black disk in the quantum well. c, Capacitance of the sample containing the artificial atom as a function of the top plate (gate) voltage. The first peak on the left represents the first electron entering the artificial atom.

dot. If the gate voltage is varied above or below these values, the number of electrons on the dot is stable and differs by one. However, when the gate voltage is set precisely at the gate voltage needed for an electron to be added to the dot, the number of electrons on the dot may fluctuate by one. This fluctuation can occur by one electron tunnelling onto the dot from the left lead and later an electron tunnelling off the dot to the right lead. This process gives rise to a detectable current flowing through the dot. In the experiment, a current peak therefore appears as the gate voltage is swept through the position where one more electron is added to the dot. These peaks are clearly demonstrated in the gate voltage scan of Fig. 2b. Unlike the SECS experiment, the first peak on the left corresponds not to the first electron in the quantum dot but to perhaps the thirtieth. To have electrical conduction through the dot, it must be close enough to the right and left leads for tunnelling to occur to both. For GTS experiments, it has as yet proven practically impossible to create a dot containing fewer than around ten electrons^{23,24}, for which appreciable electron tunnelling occurs to both electrodes. Most GTS experiments start with 25 or more electrons in the dot²⁵⁻²⁷.

As with the SECS experiment, the peak positions reflect the energy required to add each successive electron to the dot. The peaks have a non-zero width because there is a range of gate voltages over which fluctuations in the electron number on the dot can occur. Once again, this range is directly related to the temperature of the sample.

Interpreting the spectra

In both Figs 1c and 2b, electrons can then be counted one by one as they move into the artificial atom. A special feature of the SECS experiment is that it even allows precise knowledge of the number of electrons in the dot by allowing observation of the first electron in the dot.

The gate voltage scales in Figs 1c and 2b can be directly understood in terms of the energy required to add each successive electron to the quantum dot. There are essentially two reasons why it takes extra energy to add each additional electron to the dot^{12,19}. First, the electrons in the dot make it more difficult for other electrons to enter the dot. As more electrons are added, this repulsion increases and it takes more energy (charging energy) to add another electron. Second, the Pauli exclusion principle requires electrons to be in different quantum levels in the dot. Additional electrons must be promoted into higher energy levels (quantum level spacing). In the types of GaAs quantum dots discussed in this Review, the charging energies are typically a factor of 5 greater than the mean quantum level spacing.

The application of a gate voltage balances these energy requirements so that additional electrons can be added to the dot. Electrons can lower their energies by entering the quantum dot and thereby be closer to the attractive gate²⁸. Owing to the simple geometry of these structures, this energy change varies in linear proportion to the gate voltage. Multiplying the gate voltage scales in Figs 1c or 2b by a geometric factor converts this scale into an energy scale for the quantum dot. For the case of Fig. 1c, this geometric factor is 0.5. The spacings between the peaks in both figures then directly reflect the additional energies required to add successive electrons to the quantum dot. The SECS and GTS techniques thereby yield direct energy-level spectroscopy of quantum dots.

A model for the artificial atom

To understand the behaviour of the dot containing more than one electron, we need to examine in some detail the quantum energy levels of a quantum dot containing one electron. In most experiments, an electron is tightly bound in a quantum well in the z direction and has no freedom to move in this direction. I will then consider the electron to be bound laterally in the quantum well in the x - y plane by a potential $1/2m\omega_0^2r^2$. Here, $r^2 = x^2 + y^2$ measures the distance from the centre of the quantum dot. The simple mechanical analogue of this system is a ball in a bowl whose cross-section is a parabola, and ω_0 is the frequency that a classical particle would exhibit while oscillating in this bowl. This case of a circularly symmetric parabolic potential approximates well with the actual case in experiments and turns out to be rather simple to solve²⁹.

In solving the Schrödinger equation for the parabolic potential, one obtains quantum level energies given by the simple formula:

$$E_{n,l} = \hbar\omega_0(2n + |l| + 1) \quad (1)$$

where $\hbar$ is Planck's constant divided by 2π . There are two quantum numbers, n and l , corresponding to the fact that the electron moves in two dimensions. n is a positive integer which corresponds to the number of nodes in the wavefunction as one moves radially out from the dot centre, and $2|l|$ gives the number of nodes seen in moving circumferentially about the dot centre. l may be any integer, positive or negative. States of positive l correspond to electron wavefunctions moving anticlockwise with time as viewed looking down from the positive z axis, negative l states move in the opposite direction. These correspond to a series of circular orbits, with states of larger $|l|$ orbiting farther away from the dot centre. Figure 3a displays a picture of the orbits for a particular value of n . The mean radius of the different l states increases as $\sqrt{l}$.

Apart from having the quantum numbers n and l discussed above, electrons also have a spin quantum number, s , which can take values of $\pm 1/2$. An electron with $s = +1/2$ has a spin which points along the z axis perpendicular to the plane of the quantum dot (spin up); an electron with $s = -1/2$ has a spin which points

against it (spin down). In zero magnitude field, electrons with either spin state have precisely the same energy.

It is useful to examine the energy levels determined by equation (1). Plotting the energy levels as a function of l , we obtain the results of Fig. 3b. The lines in the figure connect states that have the same value of n . The bottom 'V' is for states with $n = 0$. Each state can hold two electrons, one with electron spin up and one with spin down. At low temperatures, non-interacting electrons will simply fall like marbles into the lowest available states. Each of the circles shown in Fig. 3b can hold two electrons; the filled circles represent filled states in a quantum dot containing 30 electrons. Note that for this case of zero magnetic field, there are typically several states for the same value of the energy; that is, the states are degenerate.

If a magnetic field, B , pointed in the z direction is applied, the energies and orbits of the levels change somewhat. Classically, in the absence of any confining potential, the magnetic field effectively confines electrons, causing them to move in circles in the clockwise direction with a frequency $\omega_c = eB/mc$. The quantum-mechanical energy levels are now given by:

$$E_{n,l} = [\hbar\omega_c/2 + \eta\sqrt{(\omega_c/2)^2 + \omega_0^2}(2n + |l| + 1)] \quad (2)$$

The energy has changed in two basic respects. There is now a first term that depends on l and not just the absolute value of l . Therefore, electron states with negative l values have lower energies than those with positive l values. Because the different l states represent wavefunctions that rotate in time about the dot centre, there is a magnetic moment associated with these states. This is the reason for the first term in equation (2). Positive l states have magnetic moments pointing opposite to the applied magnetic field, giving them higher energy than the negative l states whose moments point along the field. Finally, the second term in equation (2) is somewhat different from equation (1). The energies are higher, and this is because the magnetic field has enhanced the electron confinement in the dot.

The magnetic field also creates an energy difference between the spin-up and spin-down electrons in a state. The electron spins have an associated magnetic moment, and so in an externally

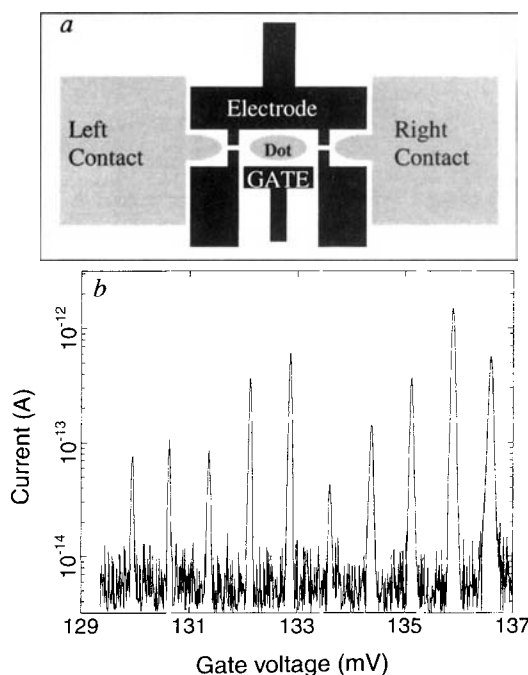


FIG. 2a, Schematic diagram of samples used in gated transport experiments. b, Current through an artificial sample as a function of gate bias. The first peak on the left is thought to represent the addition of the thirtieth electron to the artificial atom. (Courtesy O. Klein)

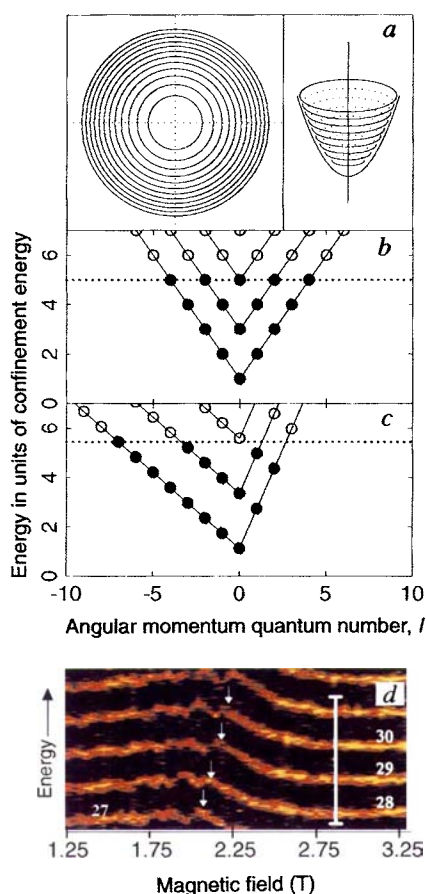


FIG. 3 **a** Left, orbits (as seen looking down on the two-dimensional artificial atom) for different values of magnitude of angular momentum $|l|$ in the artificial atom. The radius of each orbit increases as the square root of $|l|$. Right, orbits as seen in the bowl-like potential of the artificial atom. **b**, Energies of the different quantum levels as a function of angular momentum quantum number $|l|$ in zero magnetic field. The different 'Vs' correspond to different values of the radial quantum number, n . **c**, As **b**, but in the presence of a magnetic field. **d**, Colour-coded image from a SECS experiment. The red and yellow traces display the energy required to add each successive electron to the dot. To the right of the arrows, all electrons are in the lowest Landau level. The traces are numbered corresponding to the number of electrons contained in the artificial atom. The vertical bar corresponds to an energy of 5 meV.

applied magnetic field the electrons act rather like bar magnets, tending to align their moments with the external field (Zeeman effect). The energy difference between spin-up and spin-down states is often small compared to the other energy splittings in the quantum dot. To simplify the discussion, I first consider this 'spin splitting' here to be zero.

Figure 3c displays the quantum levels after the application of a magnetic field. Essentially two things have happened due to the field. The 'Vs' have rotated to the left, and the energy separation between the Vs has increased. If we again think of electrons as marbles filling slots, the marbles then fall from states on the right arm of the Vs to the left arms. Moreover, because the spacing between the Vs increases with increased field, the marbles fall from the upper to the lower Vs. In a magnetic field, we refer to each of the Vs as Landau levels. At very high fields, all of the marbles are in the left arm of the lowest V. In this case, all of the electrons are in the 'lowest Landau level'.

The peak positions in the SECS or the GTS experiment simply track the energy of the highest-energy electron in the quantum dot. This electron may be in a different quantum level depending on the magnetic field, and each of these levels has a different

evolution in a magnetic field. Therefore, the peak position is expected to zig and zag as the highest-energy electron moves from state to state. In the absence of spin splitting, each of the energy levels depicted in Fig. 3b and c holds two electrons of precisely the same energy. In this case, one would expect the two electrons to undergo identical level shifts (zigs and zags) as the magnetic field strength is varied. In reality, each level of Fig. 3c is split into two different energies (two circles displaced vertically, one for spin up and the other for spin down) by the magnetic field. At 2 tesla (2 T), the amount of this splitting is only about 0.05 meV, or approximately 0.5 K in temperature units. This value is rather small compared to the energy differences between orbitals in a quantum dot which are typically around 1–2 meV.

Both GTS²⁵ and SECS²¹ measurements demonstrate the zigzag behaviour. In the SECS experiments, one creates plots such as the one shown in Fig. 3d. This is a compendium of many data sets such as the one shown in Fig. 1c, now taken at varying values of the magnetic field strength. In Fig. 3d, the horizontal axis represents magnetic field strength and the vertical axis represents the voltage across the capacitor (or energy). The capacitance is plotted on a colour scale with yellow, red and black representing the highest, intermediate and lowest capacitances, respectively. Capacitance peaks are therefore signified by yellow and red traces. The traces clearly display a zigzag behaviour that ceases abruptly, just as we would expect from the simple-minded model of Fig. 3c. The zigzags stop at the positions of the arrows in Fig. 3d, the magnetic field strength beyond which all of the electrons are in the lowest Landau level.

Subtle differences exist between the observed traces and the results expected from the model of Fig. 3b and c. Because the spin splitting is very small, the model predicts pairs of two successive electron traces (such as for electrons 27 and 28 or electrons 29 and 30 in Fig. 3d) to have nearly identical zigzag features. This is not the case; no pairs of nearly identical traces are seen²¹. Also, the model predicts zigzags to occur well before the magnetic field value at which all electrons fall into the lowest Landau level; the data only display zigzags just below this field value. What is responsible for these and other deviations from the ideal model behaviour? One issue is most suspect: the simple model neglects the fact that the electrons interact with each other.

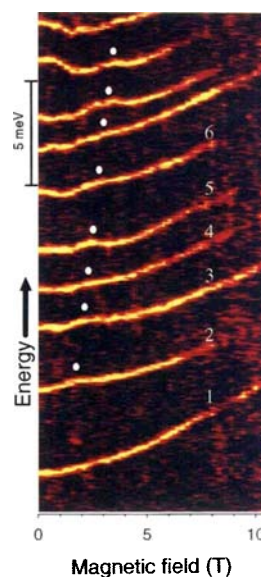


FIG. 4a, Colour-coded image from a SECS experiment for the first ten electrons in a quantum dot. The lowest red and yellow trace displays the energy required to add the first electron into the dot. The white circles indicate a kink which is thought to be the field at which all of the electronic spins align with the magnetic field.

Adding one to ten electrons

There is another important change in the quantum states with the application of magnetic field. States of all l values shrink in radius as the magnetic field is increased. For high magnetic fields (when ω_c is larger than ω_0), the radius of a particular l state shrinks as $1/\sqrt{B}$. All of the circles shown in Fig. 3a tend to converge to the dot's centre.

Let us now examine the bottom trace of Fig. 4. This trace shows SECS results for the magnetic-field dependence of the energy of the quantum dot containing only one electron. Note that this energy increases as the magnetic field strength is increased owing to the enhanced confinement of the electron by the magnetic field. This behaviour of the one-electron energy turns out to be easily predictable using equation (2). This electron stays in the lowest energy state of the quantum dot, with $n = 0$ and $l = 0$. A fit to this trace using equation (2) determines that the diameter of the first electron's wavefunction is about 40 nm.

For more than one electron in the dot, interactions between electrons make understanding the spectra substantially more complicated and rich than one might conclude from the discussion above³⁰. This is seen in the basic problem of just two electrons in a quantum dot.

The second trace from the bottom in Fig. 4 is for two electrons in the artificial atom. Notice that this trace appears qualitatively different from that of the first electron. Rather than smoothly moving up in energy, the two-electron energy shows a very clear kink at a magnetic field strength of 1.5 T. The origin of this kink, like the zigzags seen in Fig. 3d, is a shuffling of electrons between quantum states in the dot. In this case, it is the flipping of the spin of one of the two electrons within the quantum dot.

First, I consider two electrons in the dot with no Coulomb repulsion between them. Because of the Pauli exclusion principle, the two electrons must exist in quantum states that have at least one quantum number which is not the same. In zero magnetic field, the electron energy does not depend on the spin direction. These two electrons thus fall into the lowest available energy states which have quantum numbers $n = 0$, $l = 0$, and $s = \pm 1/2$. Such a state with oppositely pointing spins is known as a singlet state. At sufficiently high field, the electron which in zero field had its spin magnetic moment pointing against the applied field will, owing to the Zeeman effect, flip its spin so that the moment points along the field. When this happens, both electrons will have their spins pointing the same way, and they will have the same value of s . In this 'triplet' state, one of the electrons moves from an $l = 0$ state to an $l = 1$ state. Thus this singlet to triplet crossing arises when the Zeeman energy exceeds the energy to promote an electron to the $l = 1$ state. For the quantum dot of Fig. 4 in the absence of Coulomb repulsion, one would expect this crossing to occur at 25 T. In reality, the crossing is observed to occur at around 1.5 T. Why is this so?

The simple model of electrons as bar magnets clearly fails. When we include the Coulomb interaction between the electrons, we see that it actually drives the singlet to triplet crossing. In fact, the singlet-triplet crossing is predicted to occur even in the absence of a Zeeman interaction. The basic physics is simple, and I will approximate the actual case by considering one-electron wavefunctions. In reality, the wavefunctions change somewhat when the Coulomb interaction is added to the problem, and the electron motions are actually correlated in time to keep the electrons widely separated.

When the two electrons both have the same spatial quantum numbers (that is, $n = 0$ and $l = 0$ for both electrons), the energy of repulsion between the two electrons is large because the electron wavefunctions are identical. In contrast, when one of the two electrons moves into an $l = -1$ state, the electrons are on average farther apart, decreasing the Coulomb energy. There is a competition between the propensity for electrons to move towards the centre of the dot where the confinement potential is low and the tendency for electrons to repel each other. As the magnetic field

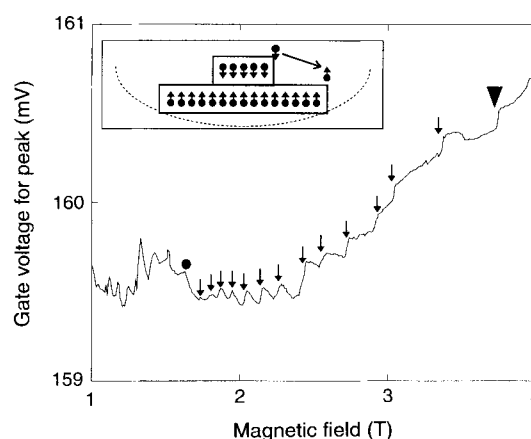


FIG. 5 Peak positions from a gated transport spectroscopy (GTS) experiment for a quantum dot containing about 30 electrons. The filled black circle marks the magnetic field value at which all of the electrons fall into the lowest orbital Landau level. The small arrows demarcate the spin flips of individual electrons in the dot. The last arrow on the right marks the field at which the electrons in the dot become spin-polarized. The black triangle marks the field values at which 'hole bunches' are created within the dot (adapted from ref. 34). Inset, schematic diagram of the spin-flipping process.

strength is increased in the two-electron dot, states of larger $|l|$ move closer to the dot centre. When the radius of the $l = -1$ state becomes sufficiently small, the balance tips in favour of one electron moving from the dot centre out to the $l = -1$ state. To satisfy the Pauli exclusion principle, the wavefunction describing the two-electron system must change sign when the coordinates of the two electrons are interchanged. This turns out to mean that an electron must flip its spin direction as it moves to the $l = -1$ state. The two-electron problem has been solved exactly^{31,32}, and the predicted singlet to triplet crossing falls close to that observed in Fig. 4.

What about the dot containing more electrons? The spectra can no longer be predicted exactly. There exist several methods for deriving approximate spectra, the most accurate of which is known as 'exact diagonalization'^{26,33–35}. This method is especially computer-intensive, and calculations can only be made for the case of a few electrons in the dot. In exact diagonalization, one typically approximates the many-body electron wavefunctions using combinations of a finite number of single-electron wavefunctions. These methods have been used to model the data in Fig. 4, and it is possible to identify the 'kinks' in these traces with predicted transitions. The main series of kinks, indicated by white circles in each of the traces, is thought to be the magnetic field value of the transition for flipping the last electron moment in line with the magnetic field³⁵. Once again, these transitions are largely driven by the Coulomb interaction between electrons and not by the interaction of the electron moments with the external field.

Many-electron interactions

For yet more electrons, there is again strong evidence that transitions in the dot are driven by interactions between the electrons. Figure 5 displays peak positions as a function of magnetic field for conductance measurements in a GTS experiment at a temperature of about 0.08 K. This dot contains around 30 electrons. The filled circle indicates the magnetic field value, as marked by the arrows in Fig. 3d, where the last electron falls into the left arm of the lowest V of Fig. 3c. These lower-temperature measurements reveal something else. Beyond the filled circle, a new series of kinks develops in the spectra. As all of the electrons are in the lowest possible spatial energy states, these kinks must be due to spin-flips. As with the two-electron case above, the Zeeman interaction is simply too weak to cause these energy level crossings.

How do the interactions produce the observed level crossings? Let us start with the case of having all of the electrons in the lowest Landau level, and each orbital l state occupied by two electrons. With the application of stronger magnetic field, these orbits (shown in Fig. 3a) tend to converge towards the centre of the dot. If no electrons were to change their orbital states, in the limit of an infinite magnetic field all of the electrons would be compressed to a point at the centre of the dot. Because electrons repel each other, this obviously cannot happen. To lower the electron density at the centre of the dot, spin-down electrons are transferred to spin-up sites at the dot's edge. This process is shown schematically in Fig. 5 inset. Even though there is only a small energy difference between spin-down and spin-up states for non-interacting electrons, the spin-down states are depopulated because there exist lower-energy states at the edge of the dot.

How does one model this situation mathematically? The simplest scheme is to account for the electron repulsion by creating a simple 'mean-field' potential which each electron feels as a result of repulsion from all of the other electrons in their average positions. This potential is added to the confinement potential in the dot, and then the electrons are allowed to rearrange into different orbits to take this potential into account. As the electrons have moved, the problem needs to be solved again for the new electron distribution. The iterations continue until the electron distribution stops changing.

This 'self-consistent' approach²⁵ does indeed produce the type of spin-flipping behaviour suggested by the data on Fig. 5. However, it fails to explain this data in detail. The spin flips in the data occur more frequently at lower fields, and less frequently at higher fields. The self-consistent model predicts a much weaker dependence on field than demonstrated in observations. What is missing here?

Electrons attract?

The self-consistent model treats the electrons as though they are independent entities whose only interaction with other electrons is through the Coulomb repulsion and through the fact that individual electrons must occupy different single-particle orbits. In reality, the interactions are much more subtle. The electron motions inside the dot are actually correlated. For instance, in the dot containing two electrons, the correlations tend to place the electrons on opposite sides of the dot. Some of this happens automatically because electrons avoid each other as a result of the Pauli exclusion principle (the fact that the many-electron wavefunction must be antisymmetric under exchange of two electrons), and this phenomenon is known as the exchange interaction. The next most sophisticated calculation scheme beyond the self-consistent method is the 'Hartree-Fock' (HF) technique, which approximates the electron repulsion using the averaged fields of all of the electrons in their orbits. However, HF does take proper account of the exchange interaction. An HF calculation, when compared with the results of Fig. 5, indeed predicts nearly the same spin-flipping behaviour³⁶.

The exchange interaction leads to a remarkable phenomenon in quantum dots. HF calculations demonstrate an apparent short-range *attraction* between electrons. Considering wavefunctions for individual electrons, the exchange interaction tends to cause electrons to fill adjacent l states without leaving any gaps. This happens because the exchange interaction is operative when single-electron wavefunctions have some spatial overlap; when the wavefunctions overlap, the many-electron wavefunction must still go to zero when the two electron positions approach each other. This phenomenon means that on average electrons in adjacent orbits are actually kept further apart (for example, at any time, they are on opposite sides of the quantum dot) than are electrons in some more widely separated orbits.

The exchange interaction produces an interesting dilemma for electrons in a quantum dot in high magnetic field. Consider the situation where the process described in Fig. 5 inset has continued as the magnetic field has been increased, so that now all of the

electrons in the dot have their spins pointing up. As the magnetic field is increased further, the orbits continue to converge to the centre of the dot. Once again, we are faced with the problem of all the electrons condensing to a point in the limit of high fields. One solution would be for electrons to leave 'holes' (empty l states) in the electron puddle—holes in the middle of the puddle which are unoccupied—so that the electrons are not spaced too closely together. However, the exchange interaction makes the creation of holes energetically very costly.

HF calculations suggest the following scenario. To save exchange energy, it is wise to keep any holes in a bunch rather than dispersing the holes throughout the dot. Figure 6b depicts the predicted behaviour. The left-hand figure shows the electron droplet with the magnetic field set just above the point at which all of the electrons become spin-polarized. The right-hand figure shows the droplet at a field just beyond a critical value at which a bunch of holes is introduced into the droplet. By putting the holes into a bunch, most of the electrons retain their nearest neighbours and hence the short-range 'attraction' is satisfied. At the same time, the size of the droplet has expanded, and the energy due to electron repulsion has been reduced. The data set displayed in Fig. 5 and the SECS results shown in Fig. 6a both demonstrate evidence of bunches of holes being admitted into the electron droplet at fields beyond the fields required for spin polarization of the droplet. Figure 6a also shows evidence of subsequent bunches being admitted as the field is increased further.

The variations in brightness of the traces in Fig. 6a arise from differences in the probability for electron tunnelling between the artificial atom and the nearby metal^{21,22}. The dim and bright regions of the traces result respectively from low and high tunnelling probabilities. One may interpret the tunnelling

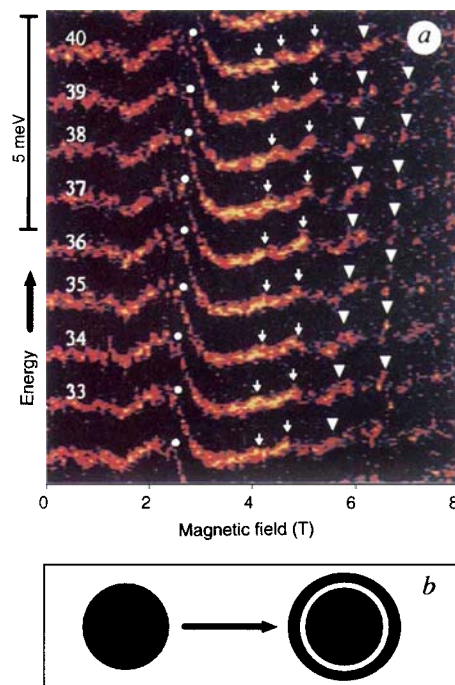


FIG. 6a, Colour-coded image from a SECS experiment. The numbers labelling each trace signify the number of electrons in the artificial atom. The white circle indicates the magnetic field value at which all of the electrons fall into the lowest Landau level. The arrows point to the spin flips of individual electrons in the dot. The last arrow on the right marks the field at which the electrons in the dot become spin-polarized. The triangles indicate the field values at which 'hole bunches' are created within the dot. b, Schematic diagram of the hole-bunch creation process. The black regions depict areas of the dot which are populated with electrons.

probability as the willingness of the electrons of the artificial atom to make room for another electron. The tunnelling probabilities are enhanced at the positions where hole bunching occurs. They are instead diminished for magnetic fields at which HF calculations predict no holes in the interior of the artificial atom, such as beyond the rightmost arrow in Fig. 6a (at around 4.5 T) when the electrons in the artificial atom become completely spin-polarized.

Although the HF calculation captures the essence of the hole-bunch creation, it ignores the effect of electron correlation on the Coulomb repulsion. Exact diagonalization and other^{37–39} calculations attempt to take these correlations into account. These suggest an even richer behaviour in the spectra as we move to higher magnetic fields and more detailed measurements.

Future research

The study of artificial atoms is a widely expanding research area. At present, the prospect of understanding electron correlations in a simple system has been a driving force for much of the theoretical work. All of the measurements discussed in this Review describe the spectroscopy of the lowest available energy levels of the artificial atoms. Measurements of excited states^{26,40,41} constitute another class of measurements which may also yield significant information on the effects of correlations^{42,43}.

An exciting new field of research is emerging. Artificial 'molecules' consisting of two or more closely spaced artificial atoms are now being created^{44,45} and studied theoretically^{46,47}. These are typically structures similar to those of Fig. 2a, but which contain more than two or more dots. In some sense, these specimens are the first step toward artificial solids made of large arrays of linked dots⁴⁸. Many of the key characteristics of real solids and real molecules, such as their magnetic and optical properties, and whether they are insulators or metals, depend largely on the strength of the couplings between the atoms. Unlike natural molecules, the strength of interactions between the neighbouring artificial atoms in artificial molecules can be varied at will (tuned) by varying the voltages on electrodes. The exciting feature here is that researchers may learn to engineer desired properties into artificial solids simply by tuning electron voltages that control the coupling between artificial atoms.

With the creation of the artificial atom, the ultimate limit of small-sized electronics is being achieved. So far, artificial atoms have allowed us to test the effects of interactions between electrons in an unprecedented way. An open question remains the practicality of using the ability to manipulate and trap single

electrons for electronic devices. A device similar to that of Fig. 2a and known as the single-electron transistor (SET) has demonstrated spectacular sensitivity for sensing of electrical charge⁴⁹, and there have been suggestions for using either SETs⁵⁰ or artificial atoms themselves⁵¹ as computing devices. Until recently, all single-electron devices functioned only at extremely low temperatures. Now, two novel structures^{2,3} for single-electron transistors have been produced in silicon, unlike the gallium arsenide structures discussed in this Review. They both clearly display room-temperature operation. The smaller a single-electron device can be produced, the larger the energy for single-electron charging, and the higher the operation temperature. It turns out that advanced silicon processing, chiefly the ability to produce controllably very thin silicon oxide layers, facilitates the production of extremely small artificial atoms.

Despite these recent improvements in the operating temperature of the devices, two important problems remain. First, it has not yet proven feasible to create devices with reproducible thresholds for adding single electrons. Present digital logic systems require sharply defined thresholds for 'on' and 'off' states of a transistor. Second, single-electron devices are inherently low-current structures, and there is a 'fan-out' problem: unlike conventional transistors, the outputs of single-electron devices cannot rapidly switch many other devices. Researchers are now learning to circumvent both of these problems. For instance, schemes have been devised for memory circuits that combine conventional transistors and single-electron devices⁵². The method is insensitive to variations in device thresholds, and hybrid designs utilizing conventional transistors can defeat the fan-out problem. Because single-electron devices are so small, huge numbers of them can be packed in the space between conventional transistors on a chip. Such an implementation might yield nearly a thousand-fold increase in the storage density of a memory chip.

The work described in this Review has shown that researchers can now routinely measure extremely small amounts of electrical charge. In fact, it is now possible to measure a charge as small as 1/10,000 of an electron charge. Theory predicts that this number might reach as low as 1/1,000,000 of an electron charge⁵³. Such a small charge might be induced on a single-electron device by a small motion of a single, distant charge. We are just now learning to make use of this extreme sensitivity. □

R. C. Ashoori is at the Department of Physics, 13-2053, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

1. Ziman, J. M. *Principles in the Theory of Solids* (Cambridge Univ. Press, 1972).
2. Yano, K. et al. *IEEE Int. Elect. Dev. Meeting in IEEE International Electron Devices Meeting* 541–544 (IEEE, New York, 1993).
3. Takahashi, Y. et al. *Electron. Lett.* **31**, 136–137 (1995).
4. Reed, M. A. et al. *Phys. Rev. Lett.* **60**, 535–537 (1988).
5. Meirav, U., Kastner, M. A. & Wind, S. J. *Phys. Rev. Lett.* **65**, 771–774 (1990).
6. Su, B., Goldman, V. I. & Cunningham, J. E. *Science* **255**, 313–315 (1992).
7. Main, P. C. et al. *Physica B* **189**, 125–134 (1993).
8. Guéret, P., Blanc, N., Germann, R. & Rothuizen, H. *Phys. Rev. Lett.* **68**, 1896–1899 (1992).
9. van der Vaart, N. C. et al. *Phys. Rev. Lett.* **73**, 320–323 (1994).
10. Schmidt, T. et al. *Phys. Rev. Lett.* **65**, 5570–5573 (1995).
11. Hansen, W. et al. *Phys. Rev. Lett.* **62**, 2168–2171 (1989).
12. Silsbee, R. H. & Ashoori, R. C. *Phys. Rev. Lett.* **64**, 1991 (1990).
13. Ashoori, R. C., Silsbee, R. H., Pfeiffer, L. N. & West, K. W. in *Nanostructures and Mesoscopic Systems* (eds Reed, M. & Kirk, W.) 323–334 (Academic, San Diego, 1992).
14. Ashoori, R. C. et al. *Phys. Rev. Lett.* **68**, 3088–3091 (1992).
15. Sikorski, C. & Merkt, U. *Phys. Rev. Lett.* **62**, 2164–2167 (1989).
16. Meurer, B., Heitmann, D. & Ploog, K. *Phys. Rev. Lett.* **68**, 1371–1374 (1992).
17. Bawendi, M. G., Steigerwald, M. L. & Brus, L. E. *Rev. Phys. Chem.* **41**, 477–496 (1990).
18. Brunner, K. et al. *Phys. Rev. Lett.* **69**, 3216–3219 (1992).
19. McEuen, P. L. et al. *Phys. Rev. Lett.* **66**, 1926–1929 (1991).
20. Cho, A. Y. *J. Cryst. Growth* **111**, 1–13 (1991).
21. Ashoori, R. C. et al. *Phys. Rev. Lett.* **71**, 613–616 (1993).
22. Ashoori, R. C. et al. *Surf. Sci.* **305**, 558–565 (1994).
23. Weis, J., Haug, R. J., von Klitzing, K. & Ploog, K. *Phys. Rev. B* **46**, 12837–12840 (1992).
24. Borsosfoldi, Z. et al. *Appl. Phys. Lett.* **66**, 3666–3668 (1995).
25. McEuen, P. L. et al. *Phys. Rev. Lett.* **64**, 11419–11422 (1992).
26. Johnson, A. T. et al. *Phys. Rev. Lett.* **69**, 1592–1595 (1992).
27. Staring, A. A. M. et al. *Phys. Rev. B* **46**, 12869–12872 (1992).
28. Kastner, M. A. *Physics Today* **46**, 24–31 (1993).
29. Darwin, C. G. *Proc. Camb. phil. Soc. Math. phys. Sci.* **27**, 86–90 (1930).
30. Maksym, P. A. & Chakraborty, T. *Phys. Rev. Lett.* **65**, 108–111 (1990).
31. Wagner, M., Merkt, U. & Chaplik, A. V. *Phys. Rev. B* **45**, 1951–1954 (1992).
32. Pfannkuche, D., Gudmundsson, V. & Maksym, P. A. *Phys. Rev. B* **47**, 2244–2250 (1993).
33. MacDonald, A. H. & Johnson, M. D. *Phys. Rev. Lett.* **70**, 3107–3110 (1993).
34. Hawrylak, P. *Phys. Rev. Lett.* **71**, 3347–3350 (1993).
35. Palacios, J. J., Martin-Moreno, L., Chiappe, G., Louis, E. & Tejedor, C. *Phys. Rev. B* **50**, 5760–5763 (1994).
36. Klein, O. et al. *Phys. Rev. Lett.* **74**, 785–788 (1995).
37. Jain, J. K. & Kawamura, T. *Europhys. Lett.* **29**, 321–326 (1995).
38. Macucci, M., Hess, K. & Iafrate, G. J. *Phys. Rev. B* **48**, 17354–17363 (1993).
39. Ferconi, M. & Vignale, G. *Phys. Rev. B* **50**, 14722–14725 (1994).
40. Foxman, E. B. et al. *Phys. Rev. B* **47**, 10020–10023 (1993).
41. Weis, J., Haug, R. J., v. Klitzing, K. & Ploog, K. *Phys. Rev. Lett.* **71**, 4019–4022 (1993).
42. Meir, Y., Wingreen, N. S. & Lee, P. A. *Phys. Rev. Lett.* **70**, 2601–2604 (1993).
43. Oaknin, J. H., Martin-Moreno, L., Palacios, J. J. & Tejedor, C. *Phys. Rev. Lett.* **74**, 5120–5123 (1995).
44. van der Vaart, N. C. et al. *Phys. Rev. Lett.* **74**, 4702–4705 (1995).
45. Waugh, S. R., Westervelt, R. M., Campman, K. & Gossard, A. C. *Phys. Rev. Lett.* **75**, 705–708 (1995).
46. Palacios, J. J. & Hawrylak, P. *Phys. Rev. B* **51**, 1769–1777 (1995).
47. Matveev, K. A., Glazman, L. I. & Baranger, H. U. *Phys. Rev. Lett.* (submitted).
48. Stafford, C. A. & Das Sarma, S. *Phys. Rev. Lett.* **72**, 3590–3593 (1994).
49. Averin, D. V. & Likharev, K. K. in *Mesoscopic Phenomena in Solids* (eds Altshuler, B. L., Lee, P. A. & Webb, R. A.) 173–271 (North-Holland, Amsterdam, 1991).
50. Likharev, K. K. *IBM J. Res. Dev.* **32**, 1444–1458 (1988).
51. Tougaw, P. D. & Lent, C. S. *J. appl. Phys.* **75**, 1818–1825 (1994).
52. Likharev, K. K. & Korotkov, A. N. in *ISDRS'95* (IEEE, New York, in the press).
53. Averin, D. V. & Likharev, K. K. in *Single Charge Tunneling* (eds Grabert, H. & Devoret, M. H.) 311–332 (Plenum, New York, 1992).

ACKNOWLEDGEMENTS. I am grateful to my collaborators at AT&T Bell Laboratories together with whom I developed SECS measurements during the time I spent there: H. Stormer, J. Weiner, L. Pfeiffer, K. Baldwin and K. West. I thank M. Bawendi and M. Kastner for their comments on this manuscript. The work described here on SECS measurements is supported primarily by the US Office of Naval Research and the David and Lucille Packard Foundation.

Protein folding in the central cavity of the GroEL–GroES chaperonin complex

Mark Mayhew, Ana C. R. da Silva, Jörg Martin*, Hediye Erdjument-Bromage†, Paul Tempst† & F. Ulrich Hartl‡

Cellular Biochemistry and Biophysics Program and Howard Hughes Medical Institute, and † Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

The chaperonin GroEL is able to mediate protein folding in its central cavity. GroEL-bound dihydrofolate reductase assumes its native conformation when the GroES cofactor caps one end of the GroEL cylinder, thereby discharging the unfolded polypeptide into an enclosed cage. Folded dihydrofolate reductase emerges upon ATP-dependent GroES release. Other proteins, such as rhodanese, may leave GroEL after having attained a conformation that is committed to fold. Incompletely folded polypeptide rebinds to GroEL, resulting in structural rearrangement for another folding trial in the chaperonin cavity.

THE chaperonins¹ mediate the ATP-dependent folding of newly synthesized polypeptides^{2–4}. GroEL, the chaperonin of *Escherichia coli*, consists of 14 subunits of relative molecular mass 57,000 (M_r 57K) which are arranged in two heptameric rings stacked back to back. The central cavity of the cylinder accepts unfolded substrate polypeptide in the conformation of a collapsed intermediate^{5–12}. Crystallographic analysis has revealed a three-domain structure of the GroEL subunits¹³. The equatorial domain contains the nucleotide-binding site and is responsible for most of the intersubunit contacts as well as the contacts between the two rings. The flexible apical domain forms the opening of the cylinder and exposes hydrophobic residues towards the cavity that might interact with hydrophobic surfaces of the substrate polypeptide^{8,14}. A small hinge domain connects the apical and equatorial domains. GroEL functionally cooperates with GroES, a single heptameric ring of ~10K subunits that binds asymmetrically to GroEL, capping one opening of the cylinder^{5,15–17}. GroES coordinates the ATP hydrolysis by GroEL with productive folding⁹, but its exact role in the chaperonin mechanism is not yet understood. The GroEL–GroES cycle^{17–21} involves the initial binding of ATP to the subunits of one GroEL toroid, which will then bind GroES^{20,22}. GroES binding is accompanied by the cooperative hydrolysis of seven ATP molecules^{18,19,23}, resulting in a stable GroEL–7ADP–GroES complex^{17–20}. Subsequent ATP hydrolysis in the GroES-distal toroid of GroEL is coupled to the exchange of the seven ADP molecules and to the transient release of GroES^{17–20}.

Substrate polypeptide interacts with GroEL through multiple ATPase cycles of GroES binding and release^{9,19,24}, suggesting that the native state is reached in iterative rounds of folding and structural rearrangement. It has been proposed that this process may occur to a significant extent in the protected environment of the GroEL cavity, thereby effectively avoiding protein aggregation^{5,6,18,25–27}. In contrast, alternative models have suggested that the main function of GroEL is to unfold unproductive folding intermediates and then to eject them with each round of ATP hydrolysis into the bulk solution for spontaneous folding^{19,24,28}. Thus, folding would occur during jumps between GroEL molecules. Our results show that GroEL can mediate folding without a mechanistic requirement for the discharge of actively unfolded protein into free solution: dihydrofolate reductase (DHFR) folds to its native state while being transiently enclosed

in the GroEL cavity by GroES, whereas rhodanese leaves the cavity after having reached a conformation that is committed to complete folding without further chaperonin interaction. These are examples of two different types of behaviour that may be found more generally.

Folding of GroEL-crosslinked DHFR

To test whether a protein can fold without complete release from GroEL into free solution, an unfolded polypeptide was covalently tethered to the inner wall of the chaperonin cavity. Mouse DHFR was chosen as a model protein⁹ because its folding can be monitored by the binding of methotrexate (MTX), a folate antagonist that interacts only with the native state and with a native-like folding intermediate of DHFR^{29,30}. Though not an *in vivo* substrate of GroEL, mouse DHFR uses the GroEL homologue Hsp60 for folding after import into intact mitochondria^{31,32}. Unlike rhodanese, the chaperonin-assisted folding of DHFR *in vitro* can occur independently of GroES but is accelerated by the cofactor⁹. As shown below, this property helped to reveal important aspects of the GroES mechanism. The crosslinking experiments were carried out with a fusion protein, DHFR–C-6His (M_r 23K), in which DHFR carries a C-terminal extension containing a trypsin cleavage site and a unique cysteine residue followed by a 6-histidine tag (Fig. 1a). The heterobifunctional crosslinker ASIB, which carries an SH-reactive iodoacetyl group and a light-sensitive azido group, was attached to the single cysteine in DHFR–C-6His after unfolding the protein in guanidinium chloride (GdmCl). The denatured protein bound efficiently to GroEL and was isolated as a complex with the ~800K chaperonin by size-exclusion chromatography. Subsequent exposure to ultraviolet light efficiently produced a major crosslinking product of apparent M_r 91K and a minor product of 86K (Fig. 1b, lane 4), suggesting that crosslinking occurred to different sites in GroEL.

To identify these attachment sites(s), GroEL-crosslinked DHFR–C-6His was digested with trypsin and the peptide(s) of interest isolated on Ni²⁺-NTA agarose. N-terminal sequencing revealed one sequence from DHFR–C-6His and one sequence from GroEL in approximately equimolar amounts, probably derived from the major crosslinking product (Fig. 1c). The sequence matching DHFR–C-6His started C-terminal of the engineered trypsin cleavage site and included a gap at position 194, corresponding to the cysteine carrying the crosslinker. The GroEL sequence spanned residues 198–206. This region contains several residues that when mutated abolish polypeptide binding to GroEL¹⁴. One of them, tryptophan 203, located at the lower edge of the apical GroEL-domain-facing the cavity¹³, was not detectable

* Present address: Department of Molecular and Cell Biology and Biochemistry, Brown University, Box G-J2, Providence, Rhode Island 02912, USA.

‡ To whom correspondence should be addressed.

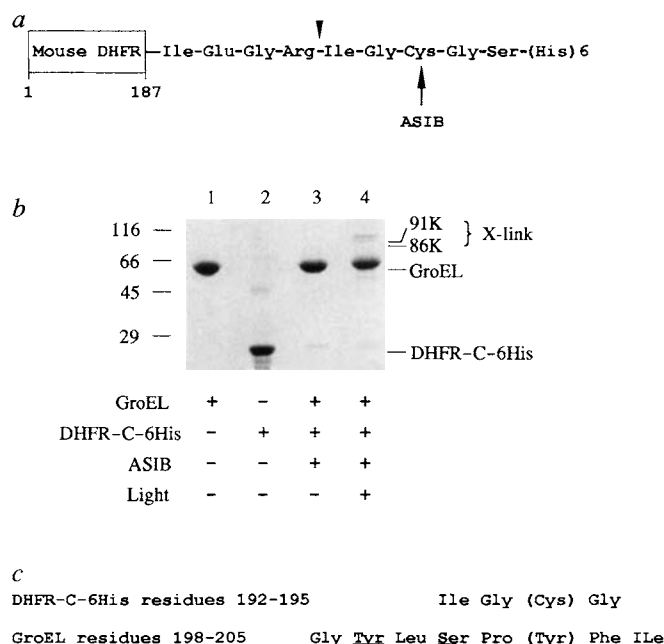


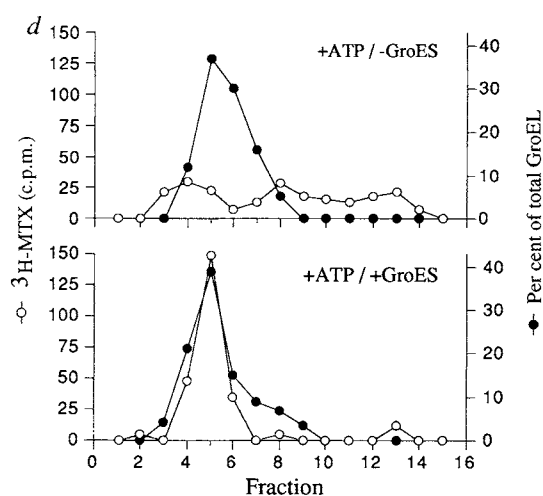
FIG. 1 Crosslinking of DHFR to GroEL and folding of crosslinked protein. **a**, Schematic representation of the DHFR-C-6His fusion protein. Arrow-head indicates a trypsin cleavage site. **b**, Incubation of GroEL and crosslinker-modified DHFR-C-6His with and without exposure to UV light followed by SDS-PAGE. Lanes 1 and 2, 5 μ g GroEL and DHFR-C-6His, respectively; lanes 3 and 4, 5 μ g GroEL: DHFR-C-6His complex. **c**, Sequences of the peptides from the major crosslinking site of DHFR-C-6His on GroEL. **d**, 3 H-MTX binding to complexes of GroEL with crosslinked DHFR protein (XL-DHFR-C-6His) in the presence of ATP (top) or ATP and GroES (bottom). Amounts of GroEL are shown as a percentage of the total eluted protein. A constant background of 3 H-MTX counts was subtracted. **METHODS:** **a**, DHFR-C-6His was derived from Cys 7 $\rightarrow$ Ser (ref. 42) mouse DHFR subcloned into pET3a (Novagen), containing a linker coding for the amino-acid sequence shown. DHFR-C-6His was overexpressed in *E. coli* in soluble form and purified on Ni²⁺-NTA-agarose (Qiagen). **b**, DHFR-C-6His (50–100 μ M) was unfolded at 25 $^{\circ}$ C for 30 min in 4 M GdmCl containing 50 mM MOPS, pH 8.0, 50 mM KCl and 0.5 mM dithiothreitol (DTT). After removal of DTT, the sample was incubated in the dark with 0.4 mM ASIB (1-[*p*-azido-salicylamido]-4-[iodoacetamido]butane; Pierce) at 25 $^{\circ}$ C for 3 h. Excess ASIB was removed by gel filtration. ASIB-labelled DHFR-C-6His (2 nmol) was then bound to 2 nmol GroEL 14-mer¹⁰ by 40-fold dilution.

and thus represents the likely crosslinking site for the azido group of ASIB.

The GroEL-crosslinked protein (henceforth termed XL-DHFR-C-6His) maintained an unfolded state which did not bind 3 H-labelled MTX (not shown). To analyse its ability to fold in association with GroEL, XL-DHFR-C-6His was incubated with 3 H-MTX and Mg-ATP. 3 H-MTX binding was not observed even when analysed on a size-exclusion column using a running buffer containing a constant concentration of 3 H-MTX³³ (Fig. 1d, top panel). Notably, 3 H-MTX binding to XL-DHFR-C-6His was detected when GroES was included in the reaction together with ATP (Fig. 1d, bottom panel). Approximately 10% of the total XL-DHFR-C-6His bound 3 H-MTX under conditions in which the chaperonin actively hydrolysed ATP. The observation of GroES-dependent folding of the crosslinked protein strongly supports the proposal⁹ that GroES is required for protein folding in association with GroEL. The GroES-independent folding reported for DHFR occurs at least partially in free solution, as it is inhibited by increasing the concentration of free GroEL⁹.

Closing the GroEL cage

GroES-dependent folding of XL-DHFR-C-6His was observed in the presence of ATP and ADP (Fig. 4). Both nucleotides support



Any non-bound DHFR-C-6His was removed by gel filtration on Sephacryl S-300 in buffer A (50 mM MOPS, pH 7.4, 50 mM KCl, 5 mM magnesium acetate). Crosslinking of GroEL-bound DHFR-C-6His (12 μ M; 15–20% efficiency) was by exposure to UV light (310–340 nm) for 5 min at 25 $^{\circ}$ C. **c**, XL-DHFR-C-6His was produced as in **b** using 35 S-labelled DHFR-C-6His⁴³. The GroEL complex was then dissociated into monomers in 4 M urea in buffer A⁴⁴, and the crosslinking product isolated on Ni²⁺-NTA-agarose. After washing with 50 mM MOPS, pH 8.0, 300 mM NaCl, 25 mM imidazole, specific elution was with buffer/300 mM imidazole. After exchange into 100 mM NH₄HCO₃, pH 8.0, 50 mM KCl the crosslinking product was incubated with trypsin (trypsin:substrate 1:30 w/w) for 3 h at 37 $^{\circ}$ C. The peptide(s) of interest was re-isolated on Ni²⁺-NTA-agarose as described. After separation by Tris-tricine SDS-PAGE⁴⁵ and electrotransfer onto a PVDF membrane in CAPS buffer (pH 11)/methanol (9/1 v/v), the 35 S-labelled peptide was N-terminally sequenced (Applied Biosystems 477A/120A)⁴⁶. **d**, XL-DHFR-C-6His was made as in **b**. Non-crosslinked DHFR-C-6His was removed by re-isolation on Sephacryl S-300 HR (17 $\times$ 0.7 cm column) in buffer A/1 mM ATP. ATP was then removed from GroEL-containing fractions by incubation with 10 mM *trans*-1,2-diaminocyclohexane-*N,N,N',N'*-tetraacetic acid (CDTA) followed by re-isolation on S-300. 3 H-MTX (100 nM, 90 c.p.m. per pmol; NEN) was then added to the final complex (6 μ M GroEL, $\sim$ 0.9 μ M crosslinked DHFR-C-6His). The reaction was divided into three parts and further incubated for 2 min without ATP (not shown) or with 1 mM ATP (both panels) and 12.5 μ M GroES (bottom panel). The reactions were then separated on S-300 columns (9 $\times$ 0.5 cm) equilibrated with buffer A, 3 H-MTX, ATP and 5 μ M GroES where indicated. 150- μ l fractions were analysed by scintillation counting or SDS-PAGE and Coomassie-blue staining followed by densitometry.

the formation of a GroEL-GroES complex, but only in the presence of ADP is this complex stable^{16–18,34}. It seemed possible that the association of GroES with the DHFR-bound ring of GroEL might displace the substrate from its non-covalent attachment sites on GroEL, thereby allowing it to fold while being encapsulated in a cage-like structure. If this is the case, non-crosslinked DHFR should also be retained within the GroEL-GroES complex (there is apparently no free passage between the two GroEL toroids^{13,25}). We tested this hypothesis by binding DHFR-C-6His (without crosslinker attached) to GroEL at close to 1:1 stoichiometry. In the presence of ADP, this protein remained associated with GroEL in an unfolded state that did not bind 3 H-MTX (Fig. 2a, top panel). Strikingly, when both ADP and GroES were added to the reaction, efficient binding of 3 H-MTX to the chaperonin complex was observed (Fig. 2a, middle panel). Approximately 15–20% of the total DHFR-C-6His associated with GroEL bound 3 H-MTX. The 3 H-MTX-DHFR complex was stable as the ligand was not lost during size-exclusion chromatography in the absence of free 3 H-MTX.

The folded DHFR protein retained by GroEL was possibly enclosed within the enlarged GroEL cavity underneath GroES⁸. To corroborate this, we utilized the observation that GroES prevents the access of proteinase K into the cavity of the GroEL

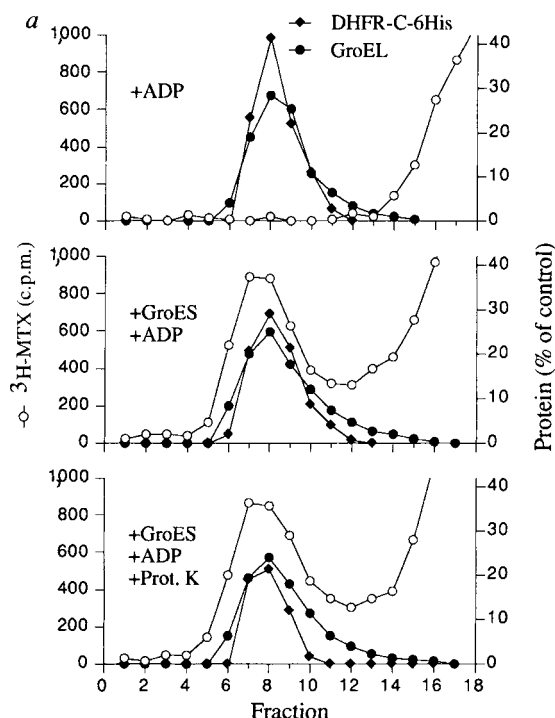
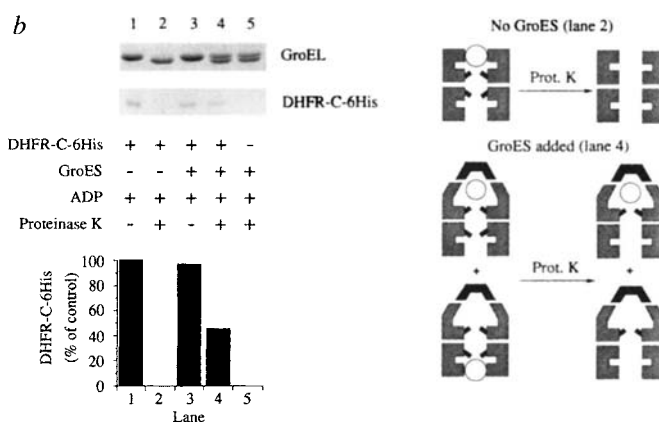


FIG. 2 Folding of non-crosslinked DHFR-C-6His in a ternary complex with GroEL and GroES. **a**, DHFR-C-6His complexes incubated in the absence (top panel) and presence of GroES (middle and bottom panels) were analysed by size-exclusion chromatography after incubation with a molar excess of ^3H -MTX. The reaction shown in the bottom panel was incubated with proteinase K before column fractionation. The profile of GroEL and DHFR-C-6His is shown as a percentage of the protein eluted in the control (reaction in top panel) and ^3H -MTX in c.p.m. per fraction. **b**, Analysis by SDS-PAGE of the effect of proteinase-K treatment on the complexes formed in **a**. The experimental situation is shown schematically for DHFR-C-6His bound to GroEL (lane 2) and for DHFR-C-6His bound in



the GroEL-ADP-GroES complex (lane 4). Unfolded DHFR-C-6His is shown as a sphere. The C-terminal segments of GroEL are shown in black (see also Fig. 6). The data in the bar diagram are averages of three independent experiments.

METHODS. **a**, GroEL:DHFR-C-6His complex was generated by incubating $5.25\ \mu\text{M}$ GroEL in buffer A with $10\ \mu\text{M}$ native DHFR-C-6His for 10 min at 45°C , resulting in unfolding and binding of DHFR-C-6His to GroEL at a stoichiometry of $\sim 0.8:1$. After cooling to 25°C , protein aggregates were removed by centrifugation (for 10 min at $10,000g$). Non-bound DHFR-C-6His was removed by isolating the chaperonin complex on a S-300 column in buffer A. The GroEL-containing fractions were concentrated to $2\ \mu\text{M}$ GroEL. The reaction was divided into three parts. When indicated, GroES was added to $4\ \text{mM}$ and all reactions received $1\ \text{mM}$ ADP and $3\ \mu\text{M}$ ^3H -MTX ($200\ \text{c.p.m. per pmol}$). For the reaction shown in the bottom panel, proteinase K was added to $10\ \mu\text{g ml}^{-1}$ and protease action stopped by addition of $1\ \text{mM}$ phenylmethylsulphonyl fluoride (PMSF) after incubation for 1 min at 25°C . All reactions were subsequently analysed by size-exclusion chromatography on a S-300 column ($17 \times 0.7\ \text{cm}$) equilibrated in buffer A. Fractions of $290\ \mu\text{l}$ were collected and analysed. **b**, Chaperonin complexes were treated with proteinase K using the protein composition indicated. Different reactions were analysed by SDS-PAGE and Coomassie-blue staining. The amount of GroEL-bound DHFR-C-6His protein present in the GroEL complex before proteinase K treatment is set to 100%.

toroid to which it is bound^{17,18}. This protects the flexible C termini of GroEL against cleavage^{5,13} (see scheme in Fig. 2b). The unfolded substrate protein bound in the opening of the cylinder is itself easily digested⁹, allowing the protease to access the C-terminal GroEL segments located below (Fig. 2b, lanes 1 and 2). If the folded and thus protease-stable DHFR⁹ in the complex were located in the GroEL toroid in a *trans* position to GroES, some protease protection of these GroEL subunits might be expected. However, when GroES and ADP were present, only the GroES-bound ring of GroEL was protected, resulting in a characteristic double-band pattern of GroEL upon SDS-PAGE^{5,17,18} (Fig. 2b, lanes 3 and 4). Interestingly, only half of the DHFR-C-6His on GroEL-GroES was digested (Fig. 2b, lane 4), without reducing the folded DHFR associated with GroEL (Fig. 2a, bottom panel). From the amount of bound ^3H -MTX, 30–40% of the DHFR-C-6His remaining after proteinase K treatment was folded. Thus, GroES enclosed $\sim 50\%$ of the GroEL-bound DHFR-C-6His in the cage-like cavity of GroEL, of which at least one third folded. The protein bound in *trans* to GroES was retained in an unfolded state and was digested by protease.

The tight binding of ^3H -MTX to the DHFR-C6-His enclosed by GroES suggested that the protein had reached its fully native conformation. This was confirmed by measuring the enzyme activity of authentic DHFR (Fig. 3a) or DHFR-C-6His in a complex with GroEL and GroES, formed in the presence of ADP. The isolated complex contained $\sim 20\%$ of the enzyme activity obtained after reactivation of all GroEL-bound DHFR upon addition of ATP (Fig. 3a, reactions 1 and 3). The same activity was measured after proteolytic digestion of the DHFR

bound in *trans* to GroES (Fig. 3a, reaction 2). Re-isolation of this complex after enzyme assay revealed that the DHFR protein in the reaction was still bound to GroEL in a position capped by GroES (Fig. 3b). Identical results were obtained with DHFR-C-6His (not shown). Notably, no lag phase was observed between the addition of enzyme substrates to the complex and the detection of enzymatic activity, in contrast to a reaction in which folding of GroEL-bound DHFR was initiated by adding ATP and GroES (not shown). We conclude that the folding of DHFR within the GroEL cavity was not 'induced' by its substrates, but occurred upon GroES-dependent displacement of the unfolded protein from its attachment sites on GroEL.

Opening the GroEL cage

How is the folded but enclosed protein released from GroEL? GroES dissociates from GroEL upon hydrolysis of seven ATP molecules in the GroES-distal toroid of GroEL. This ATP hydrolysis can be limited to a single turnover by adding Mg^{2+} chelator a few seconds after ATP¹⁹, allowing the completion of the ongoing ATPase cycle but inhibiting ATP and GroES rebinding. A complex of GroEL, DHFR-C-6His and GroES was formed in the presence of ADP. A small concentration of ^3H -MTX was added to trace the folded DHFR protein (Fig. 4a). A single round of ATP hydrolysis then released GroES from the complex, allowing all the folded substrate polypeptide to escape into the bulk solution (Fig. 4b). Thus, 15–20% of the total GroEL-bound DHFR protein (30–40% of the protein in the GroES-capped toroid of GroEL) folds in a single cycle of GroES binding and release.

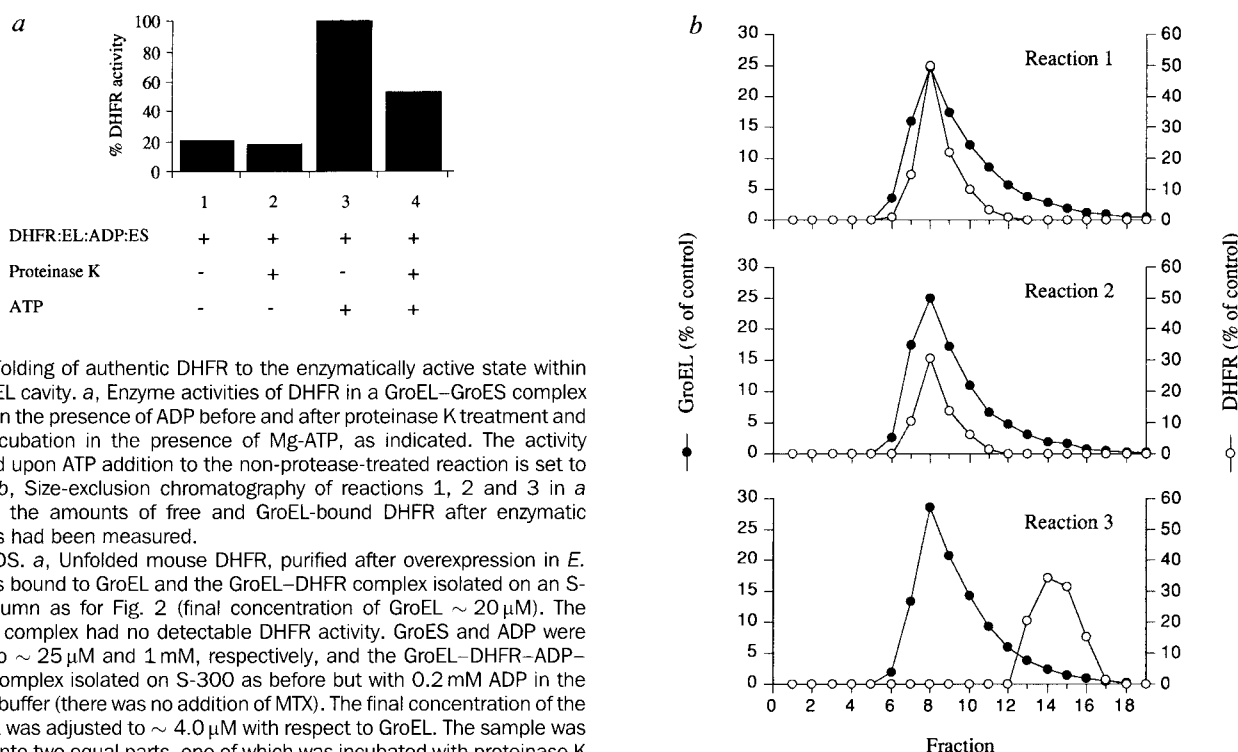


FIG. 3 Folding of authentic DHFR to the enzymatically active state within the GroEL cavity. *a*, Enzyme activities of DHFR in a GroEL–GroES complex formed in the presence of ADP before and after proteinase K treatment and upon incubation in the presence of Mg-ATP, as indicated. The activity obtained upon ATP addition to the non-protease-treated reaction is set to 100%. *b*, Size-exclusion chromatography of reactions 1, 2 and 3 in *a* showing the amounts of free and GroEL-bound DHFR after enzymatic activities had been measured.

METHODS. *a*, Unfolded mouse DHFR, purified after overexpression in *E. coli*, was bound to GroEL and the GroEL–DHFR complex isolated on an S-300 column as for Fig. 2 (final concentration of GroEL $\sim 20 \mu\text{M}$). The isolated complex had no detectable DHFR activity. GroES and ADP were added to $\sim 25 \mu\text{M}$ and 1 mM , respectively, and the GroEL–DHFR–ADP–GroES complex isolated on S-300 as before but with 0.2 mM ADP in the running buffer (there was no addition of MTX). The final concentration of the complex was adjusted to $\sim 4.0 \mu\text{M}$ with respect to GroEL. The sample was divided into two equal parts, one of which was incubated with proteinase K as for Fig. 2. Both samples were further divided into two (reactions 1 and 3, without proteinase K; reactions 2 and 4, with proteinase K). Reactions 3 and 4 received 2 mM ATP and all reactions were incubated for 20 min at 25°C . DHFR activities⁹ were determined photometrically at 25°C ($50 \mu\text{M}$ dihydrofolate and $50 \mu\text{M}$ NADPH, added from 5 mM stock solutions in buffer A). *b*, After enzyme assay, reactions 1–3 in *a* were separated by S-300

size-exclusion chromatography as for Fig. 2 in the presence of 0.2 mM ADP. The amounts of GroEL and DHFR in the column fractions were analysed by SDS–PAGE and Coomassie-blue staining followed by densitometry.

The uninterrupted GroEL/ES-mediated folding of DHFR at 25°C occurs in a first-order reaction ($t_{1/2} \sim 70 \text{ s}$; data not shown). As a single chaperonin cycle lasts $\sim 24 \text{ s}$ (refs 17, 19), this is consistent with the folding of a 20% fraction of DHFR upon each binding and release of GroES. The incompletely folded protein remaining after each cycle is recaptured by GroEL, presumably resulting in (partial) unfolding and structural rearrangement of kinetically trapped folding intermediates. By analysing the tethered protein, XL-DHFR–C-6His, the capacity of the chaperonin to unfold could be demonstrated. Although MTX binding to XL-DHFR–C-6His was readily detected in the presence of ADP and GroES (Fig. 4c), a significant fraction of XL-DHFR–C-6His, unable to leave the cavity, lost its bound MTX upon re-exposure to the polypeptide-binding regions of GroEL (Fig. 4d). Thus the chaperonin appears to have the capacity to mediate both folding and unfolding within its cavity in a manner dependent on successive rounds of GroES binding and release. This contrasts with the recent proposal that the main function of GroEL is restricted to unfolding non-native polypeptide and then discharging it into the bulk solution for spontaneous folding^{19,24}.

Folding to a committed state

The folding of rhodanese (33K) by GroEL, which is strictly GroES-dependent, has also been proposed to occur in free solution²⁴. This was concluded from results obtained with a non-cycling chaperonin mutant, GroEL337/349 (henceforth termed GroEL-trap), which was used to capture non-native folding intermediates released from wild-type GroEL. A GroEL-trap with essentially identical properties was obtained by intramolecular crosslinking of GroEL with glutaraldehyde. When either GroEL-trap was added in increasing concentrations to a pre-formed complex of GroEL and unfolded rhodanese, the yield of active enzyme measured upon addition of ATP and GroES

decreased significantly (Fig. 5a). However, a chaperonin-dependent reactivation of $\sim 20\%$ was observed even in the presence of a 20-fold molar excess of GroEL-trap over wild-type GroEL²⁴.

The residual 20% reactivation of rhodanese may either have occurred by folding in solution²⁴, or, consistent with the results described above by partial or complete folding in association with GroEL. To distinguish between these possibilities, we first measured the kinetics of rhodanese reactivation by wild-type GroEL in the presence of a 20-fold excess of GroEL-trap (Fig. 5b). Surprisingly, folding proceeded over $\sim 3 \text{ min}$. If all rhodanese molecules had to be released after every cycle of ATP hydrolysis, more than 95% trapping would have been expected. Consequently, the rhodanese molecules that folded must have maintained their association with wild-type GroEL for several rounds of ATP hydrolysis, suggesting that release of unfolded protein into free solution was not required for folding. Indeed, only 25% of the GroEL-bound rhodanese was transferred to GroEL-trap in a single ATPase cycle. This was seen when, after a single round of ATP hydrolysis, wild-type GroEL and GroEL-trap were separated by Mono-Q chromatography (Fig. 5c). Recapture of rhodanese by the opposite toroid of the same molecule of wild-type GroEL was precluded because this toroid was bound to GroES, suggesting that most of the trapped protein was released from the GroES-distal toroid of GroEL (Fig. 5c legend). These results do not support a mechanism in which all chaperonin-bound protein is ejected into the bulk solution in an unfolded state in each reaction cycle.

About 5% of the GroEL-bound rhodanese reached the native state in a single ATPase turnover (Fig. 5d, reaction 1) and about the same fraction of bound protein folded in consecutive ATPase cycles (Fig. 5b). In contrast to DHFR, folding of rhodanese to the enzymatically active state occurred only when GroES binding to GroEL was accompanied by ATP hydrolysis (Fig. 5d, reaction 2). Nevertheless, our results argue that the protein that reached the

native state underwent significant folding in the GroEL cavity. Complete folding in free solution seems highly unlikely as the yield of native enzyme produced in a single reaction cycle was unaffected when GroEL-trap was added in high excess to the preformed rhodanese-GroEL complex (reaction 3). Upon dilution from denaturant, the extended polypeptide chain is thought to collapse into a compact conformation in less than a millisecond^{35,36}. However, no rhodanese folded rapidly enough to escape GroEL binding upon dilution from denaturant (reactions 4 and 5). The spontaneous renaturation of rhodanese requires at least seconds³⁷, with the rate-limiting step occurring late in the folding pathway^{35,36}. Thus, there would be insufficient time for rhodanese to fold in bulk solution if the protein were to be released from GroEL in an extended or collapsed unfolded state. At a physiological GroEL concentration (3–5 μM), unfolded rhodanese would spend on average only ~ 3 ms in free solution before rebinding (Fig. 5d legend).

We conclude that a fraction of the bound rhodanese must leave GroEL in a folded conformation that is committed to reach the native state without further chaperonin interaction. The release of non-committed folding intermediates, measured in the trapping experiments, seems to be unrelated to the productive chaperonin mechanism. Whereas GroEL releases $\sim 5\%$ of the bound rhodanese in the committed state per cycle, an additional 25% are lost in every cycle in a non-native conformation that is unproductive for folding. Indeed, the time course of the complete folding reaction in the absence or presence of GroEL-trap could be modelled based on these parameters (Fig. 5b, dotted lines).

Discussion

Our results suggest a model in which a single chaperonin cycle can be dissected into the following steps (shown for DHFR in Fig. 6). (1) Unfolded polypeptide binds initially within the GroEL toroid in *trans* to GroES¹⁸; (2) this facilitates GroES dissociation^{17,18}; (3) GroES rebinds together with ATP to either GroEL

toroid¹⁸ (only binding to the polypeptide-containing toroid is shown). Substrate protein bound to GroEL in *cis* to GroES is displaced into the GroEL cavity and a certain fraction folds to the native or to a committed state. Some protein bound in *trans* to GroES may be released from GroEL in a non-native state not committed for folding. (4) Cooperative hydrolysis of seven ATP molecules in the *cis* toroid of GroEL stabilizes the binding of GroES. This step may be critical for the folding of certain proteins. (5) Folded (or committed) protein emerges from the GroEL cage upon GroES dissociation, triggered by ATP hydrolysis in the opposite GroEL toroid. (6) Incompletely or incorrectly folded protein rebinds tightly, resulting in unfolding and structural rearrangement. GroES rebinding then triggers another round of folding. These cycles continue until all the protein has completed folding, thus effectively preventing protein aggregation.

Is this simple model consistent with the structural information available on the chaperonin system? GroES binding causes the outwards movement of the apical domains of the interacting GroEL subunits⁸. This results in the formation of a dome-shaped chamber with approximate dimensions of 70 Å in height and width, which should in principle be sufficient to accept the collapsed intermediates of proteins up to $\sim 50\text{K}$. Recent results show that rhodanese is also efficiently enclosed in the GroEL cavity by GroES (M.K. Hayer-Hartl and F.U.H., unpublished), whereas proteins too large to be enclosed, such as β -galactosidase (M_r 116K), may fold by a GroEL-independent mechanism³⁸. GroES binding is mediated by amino-acid residues of GroEL exposed at the outer surface of the chaperonin cylinder and by the regions of the apical domains that are critical for polypeptide association^{13,14}. Thus GroES could make first contact with GroEL via the residues at the surface, and then, either directly or indirectly, displace the unfolded polypeptide from some or all of its attachment sites into the GroEL cavity^{18,39,40}.

The efficiency of folding in a single chaperonin cycle varies considerably for different proteins. Although more than 30% of

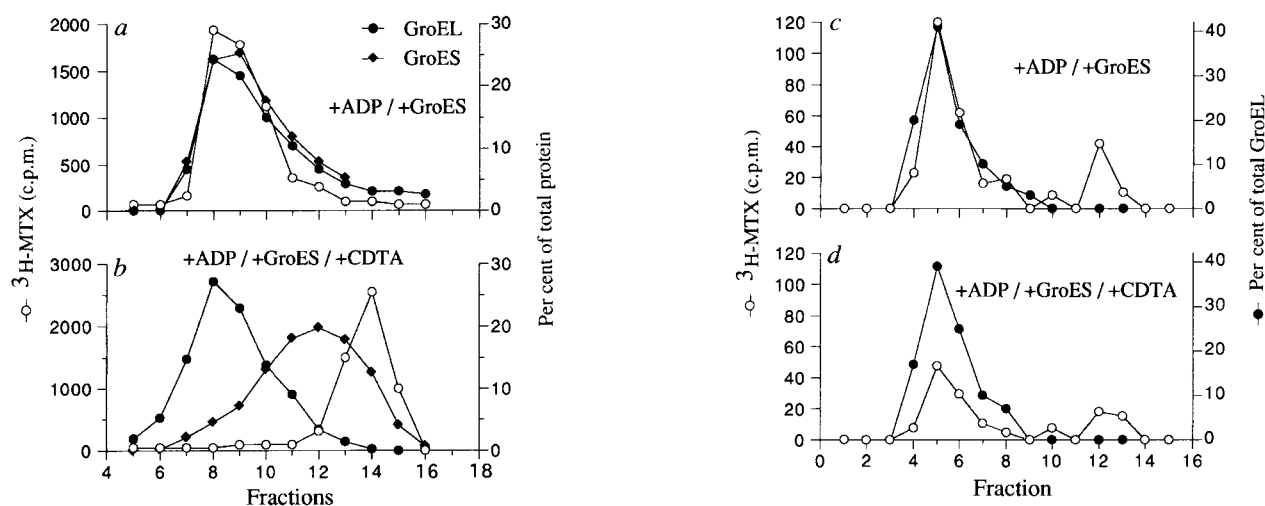


FIG. 4 Release of folded DHFR-C-6His from the GroEL-GroES complex and unfolding of DHFR-C-6His by GroEL. a, b, Non-crosslinked DHFR-C-6His bound to GroEL was folded by addition of GroES and ADP (a) and released from GroEL in a single cycle of the GroEL ATPase (b). c, d, Crosslinked DHFR-C-6His was folded in a complex with GroEL and GroES (c) and unfolded upon dissociation of GroES from GroEL (d). Amounts of GroEL and GroES in column fractions are given as a percentage of the total protein eluted from the column and ^3H -MTX in c.p.m. per fraction.

METHODS. a, b, DHFR-C-6His was bound to GroEL and the GroEL-bound protein was isolated and adjusted to 3 μM as for Fig. 2. GroES (3 μM), 0.2 mM ADP and 26 nM ^3H -MTX (3.4×10^4 c.p.m. per pmol) were added to the complex. After incubation for 2 min at 25 °C, the sample was divided

into two equal parts. ATP (2 mM) was added to one reaction followed by the addition of 10 mM CDTA after 10 s at 25 °C. Both reactions were then analysed on an S-300 column equilibrated in buffer supplemented with 1 mM ADP, essentially as for Fig. 2. ^3H -MTX counts detected beyond fraction 13 in a, derived from free ^3H -MTX, were subtracted in a and b. c, d, GroEL-crosslinked DHFR-C-6His was generated as for Fig. 1d in the presence of GroES (2-fold molar excess over GroEL), 1 mM ADP and 100 nM ^3H -MTX (90 c.p.m. per pmol). After incubation for 2 min at 25 °C, the sample was divided into two parts, one of which received 10 mM CDTA. Both reactions were then analysed by size-exclusion chromatography as for Fig. 1d with the exception that the CDTA-treated sample was separated using running buffer without Mg^{2+} .

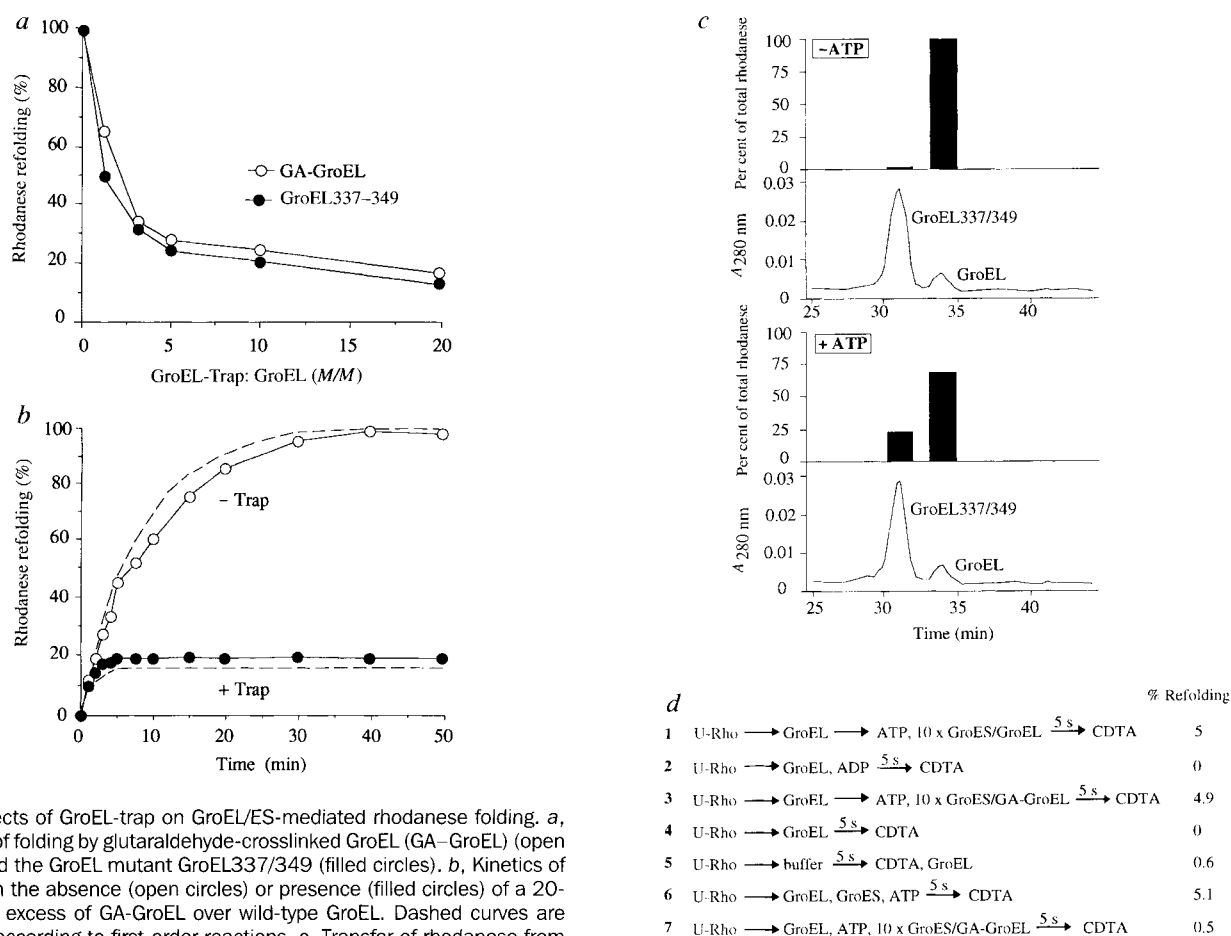


FIG. 5 Effects of GroEL-trap on GroEL/ES-mediated rhodanese folding. *a*, Inhibition of folding by glutaraldehyde-crosslinked GroEL (GA-GroEL) (open circles) and the GroEL mutant GroEL337/349 (filled circles). *b*, Kinetics of refolding in the absence (open circles) or presence (filled circles) of a 20-fold molar excess of GA-GroEL over wild-type GroEL. Dashed curves are modelled according to first-order reactions. *c*, Transfer of rhodanese from wild-type GroEL to GroEL337/349 in a single ATPase cycle detected by ion-exchange chromatography. *d*, Rhodanese refolding in a single GroEL ATPase cycle.

METHODS. GroEL337/349 (ref. 24) was constructed⁴² from the *groEL* gene. GA-GroEL was prepared by incubation of 0.4 μ M GroEL in buffer B (25 mM MOPS, pH 7.2, 75 mM KCl, 5 mM MgCl₂, 1 mM DTT) with 1.5% GA at 25 °C for 45 min. GA was removed by incubation in 40 mM sodium borohydride for 20 min, followed by exchange into buffer B. GA-GroEL did not bind GroES and ATP, but bound unfolded DHFR and rhodanese with apparently the same affinity as wild-type GroEL. *a*, Rhodanese⁴⁷ (100 μ M in 5.3 M GdmCl) was bound to GroEL (0.15 μ M) by 150-fold dilution into buffer C (buffer B plus 10 mM DTT and 50 mM sodium thiosulphate). After re-isolation, rhodanese-GroEL complexes were concentrated to 0.25 μ M GroEL and GroEL-trap was added. Refolding was initiated by adding 3 mM ATP and 6 μ M GroES. After 6 min at 25 °C, refolding was stopped with 15 mM CDTA and enzyme activities were assayed at 25 °C (ref. 9). *b*, Reactivation of rhodanese was followed by incubating a rhodanese-GroEL complex (0.25 μ M) (see *a*) in buffer C with 3 mM ATP and 6 μ M GroES either with or without 5 μ M GA-GroEL. CDTA was added at the times indicated. Modelling of first-order reactions was based on a single-round folding yield of 5%, a 24-s ATPase cycle^{17,19} and a 25% transfer of GroEL-bound protein to trap in each reaction cycle. 100% reactivation equals 80% of the activity of a native enzyme control. *c*, A complex of ³H-labelled rhodanese⁴⁷ and GroEL (0.16 μ M) was incubated in buffer C with 0.64 μ M GroEL337/349 and 0.8 μ M GroES in the absence or presence of 3 mM ATP. CDTA was

added 5 s after ATP. Samples were applied onto a 1-ml Mono-Q column (Pharmacia), equilibrated in 25 mM MOPS, pH 7.2, 50 mM NaCl, 1 mM DTT, and eluted with 50–500 mM NaCl. A_{280} and ³H-rhodanese content of the fractions were determined. Similar results were obtained if rhodanese transfer was initiated by 0.5 mM ATP and the reaction stopped by addition of either 5 mM ADP or glucose (55 mM) and hexokinase (12 U ml⁻¹; Böhringer) to maintain the GroES-bound state of GroEL. A correction factor of 3–5% was applied for the backwards transfer from GroEL-trap to wild-type GroEL²⁴. *d*, Denatured rhodanese was diluted to 0.2 μ M into buffer C containing the components indicated. Reactions 1–7 received 0.6 μ M GroEL, 1 or 6 μ M GroES, 6 μ M GA-GroEL, 2 mM ATP or ADP and 15 mM CDTA as shown. Rhodanese activities are expressed as percentage of the maximal ATP-dependent reactivation after a 60-min incubation (as in *b*). The frequency, *J*, of encounters between a molecule or rhodanese and GroEL in solution was estimated as $J = 4\pi D_{\text{GroEL}}(r_{\text{rho}} + r_{\text{GroEL}})c_{\text{GroEL}}$, with the diffusion rate of GroEL, $D_{\text{GroEL}} = 10^{-7}$ cm² s⁻¹, the radius of rhodanese, $r_{\text{rho}} = 25$ Å, the radius of GroEL, $r_{\text{GroEL}} = 70$ Å and a physiological concentration of GroEL, c_{GroEL} of 4 μ M. Assuming that binding of unfolded rhodanese to GroEL, both in the absence and presence of ATP, occurs at a rate close to diffusion-limited^{26,27} and that only 10% of the collisions result in the formation of a rhodanese-GroEL complex, a rhodanese molecule would spend 3.5 ms in solution before rebinding to GroEL.

the DHFR located underneath the hood of GroES folds in a single round, the yield of folded rhodanese is lower, about 5% of total GroEL-bound protein. We assume that folding occurs during the conformational changes in GroEL that are induced by the binding of GroES. Incompletely folded protein, being kinetically trapped, is apparently recaptured by GroEL. This step probably precedes the release of GroES because only a fraction of the enclosed DHFR folds, although DHFR is capable of spontaneous folding *in vitro*. GroES release then makes the polypeptide binding

regions of the chaperonin fully available for high-affinity rebinding of non-native polypeptide. Rebinding to GroEL, acting as a proof-reading mechanism, would lead to a more unfolded state and induce conformational changes in the polypeptide. Although only certain proteins, such as DHFR, may reach their native state enclosed within the GroEL cavity, we have demonstrated that rhodanese leaves the cavity in a conformation that is committed to proceed to the native state without rebinding to chaperonin. In the committed state the protein is likely to have

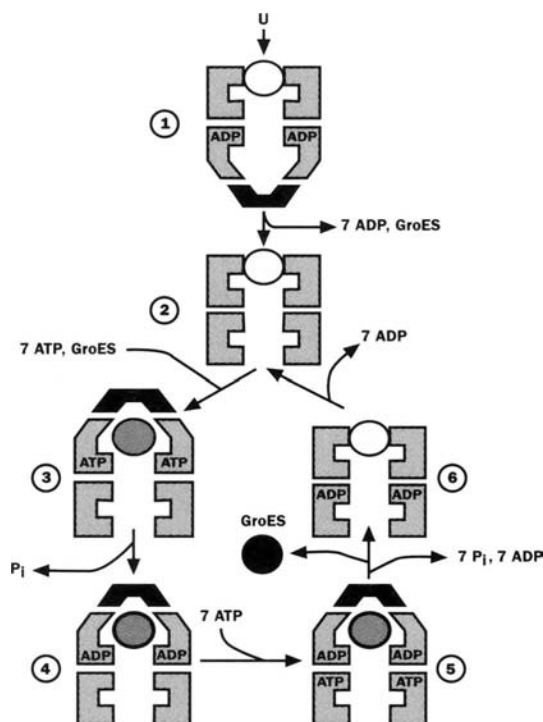


FIG. 6 Model for the chaperonin-mediated folding of dihydrofolate reductase protein. GroEL is shown schematically as a vertical cut through the complex, reflecting the three-domain structure of the subunits and the conformational changes upon GroES binding⁸. ADP, the high-affinity ADP state in the seven subunits of GroEL that are bound to GroES; ATP, the subunits in a GroEL toroid in the ATP-bound state; U, unfolded polypeptide substrate as compact folding intermediate (open sphere). The initial release of GroES from the *trans*-toroid of GroEL in (1) may be accompanied by a round of ATP hydrolysis in the polypeptide-containing toroid (not shown). Shaded spheres indicate that the population of GroEL molecules contain a mixture of folded and incompletely folded DHFR molecules. Between steps 5 and 6, folded DHFR is released (black sphere), whereas incompletely folded DHFR rebinds, resulting in structural rearrangement and unfolding. See text for details.

buried a significant fraction of its hydrophobic residues, thus allowing release from GroEL and rapid completion of folding.

Our findings do not support the view that the chaperonin mechanism in productive folding requires the ejection of actively unfolded, non-committed protein species into the bulk solution^{19,24}. Rather, the observed dissociation of such non-native intermediates reflects leakage that may occur predominantly from the GroEL-toroid in *trans* to GroES. In the case of DHFR, this protein also reaches the native state. But in contrast to the protein released from underneath GroES, these molecules leave GroEL in an unfolded, non-committed state which can be

trapped by another GroEL molecule. In the case of rhodanese, the molecules released from GroEL in the absence of GroES are prone to aggregation⁹. Interestingly, this premature release of non-native intermediate, observed *in vitro*, can be restricted by macromolecular crowding agents⁴¹ which may mimic the excluded-volume effects prevailing in the cytosol (J.M. and F.U.H., unpublished). Some release of non-native intermediate from GroEL is likely to occur physiologically to allow the degradation of damaged protein that is incapable of folding. Release from the *trans*-toroid may be the mechanism to clear these polypeptides from the chaperonin. □

Received 25 August; accepted 20 December 1995.

- Hemmingsen, S. M. et al. *Nature* **333**, 330–334 (1988).
- Hendrick, J. P. & Hartl, F. U. A. *Rev. Biochem.* **62**, 349–384 (1993).
- Ellis, R. J. *Opin. struct. Biol.* **4**, 117–122 (1994).
- Landry, S. J. & Gierasch, L. M. A. *Rev. Biophys. Biomol. Struct.* **23**, 645–669 (1994).
- Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F. U. *EMBO J.* **11**, 4757–4765 (1992).
- Creighton, T. E. *Nature* **352**, 17–18 (1991).
- Braig, K., Simon, M., Furuya, F., Hainfeld, J. F. & Horwich, A. L. *Proc. natn. Acad. Sci. U.S.A.* **90**, 3978–3982 (1993).
- Chen, S. et al. *Nature* **371**, 261–264 (1994).
- Martin, J. et al. *Nature* **352**, 36–42 (1991).
- Hayer-Hartl, M. K., Ewbank, J. J., Creighton, T. E. & Hartl, F. U. *EMBO J.* **13**, 3192–3202 (1994).
- Zahn, R., Spitzfaden, C., Ottiger, M., Wuthrich, K. & Pluckthun, A. *Nature* **368**, 261–265 (1994).
- Robinson, C. V. et al. *Nature* **372**, 646–651 (1994).
- Braig, K. et al. *Nature* **371**, 578–586 (1994).
- Fenton, W. A., Kashi, Y., Furtak, K. & Horwich, A. L. *Nature* **371**, 615–619 (1994).
- Saibil, H., Dong, Z., Wood, S. & auf der Mauer, A. *Nature* **353**, 25–26 (1991).
- Engel, A. et al. *Science* **269**, 832–836 (1995).
- Hayer-Hartl, M. K., Martin, J. & Hartl, F. U. *Science* **269**, 836–841 (1995).
- Martin, J., Mayhew, M., Langer, T. & Hartl, F. U. *Nature* **366**, 228–233 (1993).
- Todd, M. J., Viitanen, P. V. & Lorimer, G. H. *Science* **265**, 659–666 (1994).
- Burston, S. G., Ranson, N. A. & Clarke, A. R. *J. molec. Biol.* **249**, 138–152 (1995).
- Hendrick, J. P. & Hartl, F. U. *FASEB J.* **9**, 1559–1569 (1995).
- Bochkareva, E. S. & Girshovich, A. S. *J. biol. Chem.* **269**, 23869–23871 (1994).
- Gray, T. E. & Fersht, A. R. *FEBS Lett.* **292**, 254–258 (1991).
- Todd, M. J., Kashi, Y., Fenton, W. A. & Horwich, A. L. *Cell* **78**, 693–702 (1994).
- Saibil, H. R. et al. *Curr. Biol.* **3**, 265–273 (1993).
- Gray, T. E. & Fersht, A. R. *J. molec. Biol.* **232**, 1197–1207 (1993).

- Corrales, F. J. & Fersht, A. R. *Proc. natn. Acad. Sci. U.S.A.* **92**, 5326–5330 (1995).
- Ranson, N. A., Dunster, N. J., Burston, S. G. & Clarke, A. R. *J. molec. Biol.* **250**, 581–586 (1995).
- Touchette, N., Perry, K. & Matthews, C. R. *Biochemistry* **25**, 5445–5452 (1986).
- Frieden, C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4413–4416 (1990).
- Ostermann, J., Horwich, A. L., Neupert, W. & Hartl, F. U. *Nature* **341**, 125–130 (1989).
- Rassow, J. et al. *Molec. cell. Biol.* **15**, 2654–2662 (1995).
- Hummel, J. P. & Dryer, W. J. *Biochim. biophys. Acta* **63**, 530–532 (1962).
- Jackson, G. S. et al. *Biochemistry* **32**, 2554–2563 (1993).
- Creighton, T. E. *Biochem. J.* **270**, 1–16 (1990).
- Dobson, C. M., Evans, P. A. & Radford, S. E. *Trends biochem. Sci.* **19**, 31–37 (1994).
- Horowitz, P. M. & Criscimagna, N. L. *J. biol. Chem.* **265**, 2576–2583 (1990).
- Horwich, A. L., Low, K. B., Fenton, W. A., Hirschfield, I. N. & Furtak, K. *Cell* **74**, 909–917 (1993).
- Hartl, F. U. *Nature* **371**, 557–559 (1994).
- Landry, S. J., Zeilstra, R. J., Fayet, O., Georgopoulos, C. & Gierasch, L. M. *Nature* **364**, 255–258 (1993).
- Zimmerman, S. B. & Trach, S. O. *J. molec. Biol.* **222**, 599–620 (1991).
- Higuchi, R. in *PCR Protocols* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 177–183 (Academic, San Diego, CA, 1990).
- Pratt, J. M. in *Transcription and Translation: A practical approach* (eds Hames, B. D. & Higgins, S. J.) 179–210 (IRL, Oxford, 1984).
- Lissin, N. M., Vennyaminov, S. & Girshovich, A. S. *Nature* **348**, 339–342 (1990).
- Schägger, M. & von Jagow, G. *Analyt. Biochem.* **166**, 368–379 (1987).
- Tempst, P., Geromanos, S., Elicone, C. & Erdjument-Bromage, H. *Methods: A Companion to Methods in Enzymology* **6**, 248–261 (1994).
- Hlodan, R., Tempst, P. & Hartl, F. U. *Nature struct. Biol.* **2**, 587–595 (1995).

ACKNOWLEDGEMENTS. We thank R. J. Ellis for discussion and M. Hayer-Hartl for help with some of the DHFR-C-6His folding experiments. M.M. was in part supported by a HFSP postdoctoral fellowship and A.C.R.S. by a fellowship from CNPq Brazil. This work was supported by a grant from the NIH, the Howard Hughes Medical Institute and a Cancer Center core grant.

Discovery of a hierarchical distribution of dark matter in the Fornax cluster of galaxies

Y. Ikebe^{*†}, H. Ezawa^{*}, Y. Fukazawa^{*}, M. Hirayama^{*‡},
Y. Ishisaki^{*}, K. Kikuchi[§], H. Kubo^{*‡}, K. Makishima^{*},
K. Matsushita^{*}, T. Ohashi[§], T. Takahashi^{*‡}
& T. Tamura^{*}

^{*} Department of Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan

[†] Institute of Physical and Chemical Research, Hirosawa, Wako, Saitama 351-01, Japan

[§] Department of Physics, Tokyo Metropolitan University, Minamiosawa, Hachioji, Tokyo 192-03, Japan

[‡] Institute of Space and Astronautical Science, Yoshinodai, Sagami-hara, Kanagawa 229, Japan

THE mass of the Universe is widely believed to be dominated by dark matter¹. Dark matter is, by its very nature, extremely difficult to investigate, and the presence of dark matter is usually inferred indirectly through its gravitational influence on ordinary visible matter. Observations of X-ray-emitting gas associated with clusters of galaxies can help to constrain the large-scale distribution of dark matter; if the gas is in hydrostatic equilibrium with the gravitational potential of the cluster, it will trace the distribution of all matter present (dark and visible)². Here we present X-ray observations of gas in the Fornax cluster of galaxies, which show that dark matter is distributed on at least two distinct length scales: the scale of the dominant central galaxy and that of the cluster as a whole. This suggests the presence of either a single form of dark matter exhibiting hierarchical clustering (analogous to the hierarchical distribution of visible matter), or two forms of dark matter which interact—and hence cluster—through different unknown mechanisms.

X-ray-emitting diffuse plasmas of temperature $T = 10^7$ – 10^8 K are widely found in elliptical galaxies^{3–5} and clusters of galaxies^{2,6–9}. As these plasmas are thought to be hydrostatically trapped by gravity, their X-ray measurements yield direct information on the total (visible plus dark) gravitating matter in the systems. Specifically, the total gravitational potential ϕ of the system is related to the plasma density n and the plasma pressure $p = nkT$ as²

$$\nabla p + \mu m_p n \nabla \phi = 0 \quad (1)$$

where k , $\mu \approx 0.6$ and m_p are the Boltzmann constant, mean molecular weight and proton mass, respectively. This equation can be combined with Newton's equation of gravity, $4\pi G\rho = \nabla^2\phi$, where G is the gravitational constant and ρ is the density of the total gravitating mass including the dark matter contribution. We thus obtain

$$\rho = -\frac{k}{4\pi G\mu m_p} \nabla \left[\frac{1}{n} \nabla(nT) \right] \quad (2)$$

In addition, the hot plasmas themselves normally constitute the dominant known mass component of clusters², so that their contribution must be accurately subtracted from the value of ρ calculated from equation (2), to single out the dark matter distribution. These facts make X-ray observations of galaxies and clusters particularly important in the quest to identify dark matter.

The distribution of dark matter has therefore been measured with X-rays in many elliptical galaxies and clusters. Nevertheless, little is known of how its cluster-wide concentration is related to the distribution around each member of the galaxy cluster. For

example, a dominant elliptical galaxy (a 'cD' galaxy) is frequently found at the centre of a cluster, where the X-ray brightness often shows a local excess suggestive of the presence of an additional potential 'dimple'. In such cases, however, the plasma temperature usually decreases towards the cD galaxy, and the associated uncertainty in the temperature structure often hampers unambiguous calculations of equation (2) from galaxy to cluster scales.

An ideal opportunity to study the relation between galaxy dark matter and cluster dark matter is provided by the Fornax cluster of galaxies, which is a poor southern cluster hosting a cD galaxy, NGC1399. The galaxy velocity dispersion of the Fornax cluster, 307–325 km s⁻¹ (ref. 10), is close to the stellar velocity dispersion of NGC1399, 200–365 km s⁻¹ (ref. 11). We thus expect the galaxy potential and the cluster potential to have similar virial temperatures, which obviates the necessity to resolve the complex temperature structure. There is evidence for X-ray emission with similar temperatures ($\sim 10^7$ K) both from the local region of NGC1399 (refs 12–15) and on the cluster-wide ($\sim 1^\circ$) scale^{13–17}.

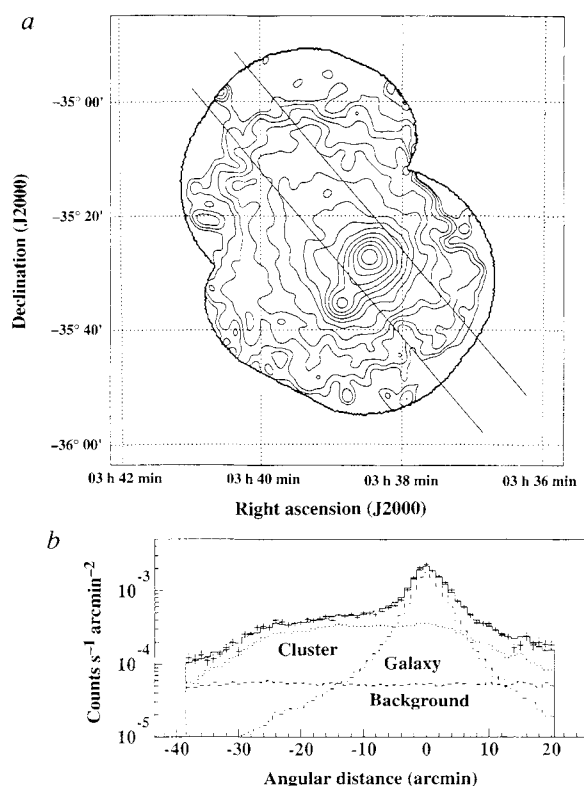


FIG. 1 a, Mosaic X-ray image of the central region of the Fornax cluster of galaxies, synthesized from three partially overlapping pointings with ASCA and shown in J2000 coordinates. Data taken at 0.7–10 keV by the two GIS detectors are added together, and the background is subtracted. The central galaxy NGC1399 appears as the brightest peak, and a companion elliptical galaxy NGC1404 forms a smaller peak to the southeast. Fainter X-ray emission is seen to extend over the entire image. The contours are logarithmic where each step corresponds to a multiplicative factor of 1.3. b, One-dimensional X-ray brightness profile in the 0.7–2.0 keV range (crosses), derived along the narrow strip in a. Histograms represent a fit with a two-component β -model defined as $\varepsilon(\mathbf{R}) = \varepsilon_g [1 + (\mathbf{R}/a_g)^2]^{-3/2} + \varepsilon_c [1 + (\mathbf{R} - \mathbf{b})^2/a_c^2]^{-3/2}$. Here $\mathbf{R}$ is the three-dimensional position measured from NGC1399, ε_g and ε_c are central emissivities, a_g and a_c are core radii, β_g and β_c are the usual β parameters, and $\mathbf{b}$ represents the relative offset between the two components. $\mathbf{b}$ is assumed to have only one degree of freedom, lying along the strip in the northeast–southwest direction within the plane of the sky. The background was not subtracted from the data but added to the model. Best-fit parameters are summarized in Table 1.

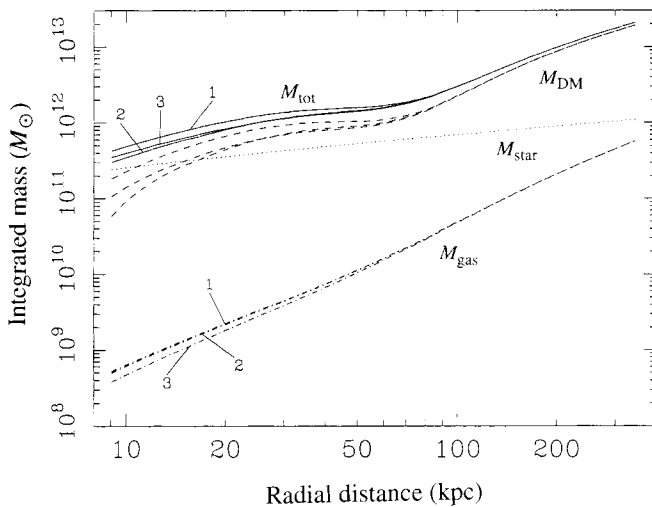


FIG. 2 Integrated mass profiles of total gravitating matter (M_{tot} , solid lines), dark matter (M_{DM} , dashed lines) and X-ray emitting plasma (M_{gas} , dot-dashed lines). Three curves for each component, specified by numbers (1–3), correspond to the three models for the plasma temperature and abundance distribution (see text for detail); model 1 is isothermal; model 2 has a central temperature decrease; model 3 has a central abundance increase. The profile of light-emitting matter (M_{star} , dotted line) was derived from ref. 12 assuming a mass-to-light ratio of 8. The integration radius is measured from NGC1399, assuming a source distance of 20 Mpc.

We accordingly observed the Fornax cluster with the X-ray satellite ASCA¹⁸, which is ideal for this purpose because of the high throughput of its X-ray telescopes (XRTs)¹⁹ and the very low background of its focal plane instruments. We made three partially overlapping pointings, with a total exposure time of 50 ks; two of them were achieved on 15 July 1993 during performance verification, the other on 25 January 1994.

Figure 1a shows a mosaic image synthesized from data taken with the Gas Imaging Spectrometer (GIS) instrument, which has a wide ($\sim 45'$ diameter) field of view¹⁸. As expected, NGC1399 appears as a clear peak, and a fainter emission is seen to extend over the entire image. This dual structure can be seen more clearly in the one-dimensional X-ray brightness profile (Fig. 1b), derived along a narrow strip in Fig. 1a. The profile appears to involve two distinct emission components (with some offset between them), one distributed over the entire cluster whereas the other is localized near NGC1399. This could, however, be due to an instrumental artefact, because the XRT response¹⁹ has wide scattering wings and complicated dependencies on position and X-ray energy.

To quantify the brightness profile of Fig. 1b, taking all the instrumental responses into account, we have developed a 'for-

ward method' analysis technique²⁰ which begins by modelling the three-dimensional distribution of emissivity $\varepsilon(\mathbf{R}) = n^2 \Lambda(T, A)$ in terms of a small number of model parameters. Here $\mathbf{R}$ is the three-dimensional position measured from NGC1399, Λ is the 'cooling function', and A is the abundance of heavy elements in solar units. We then generate Monte Carlo photons as specified by $\varepsilon(\mathbf{R})$, project them onto the sky plane and put them through the known responses of the XRT and the GIS, to obtain about a million simulated GIS events which are to be compared with the actual data. We specifically model the emissivity distribution in terms of a sum of two empirical β -models², as given in Fig. 1b, where a narrower component and a wider one are involved. By adjusting the model parameters, we have obtained an acceptable ($\chi^2/\nu = 52.3/53$) fit to the observed one-dimensional brightness profile as shown in Fig. 1b. The two β components both turn out to be significant, proving that the two-component structure seen in Fig. 1 is real. The best-fit model parameters are given in Table 1.

The best-fit emissivity ε thus obtained can be converted into a density distribution through the equation

$$n(\mathbf{R}) = \sqrt{\varepsilon(\mathbf{R})/\Lambda(T, A)} \quad (3)$$

which can then be substituted back into equation (2) to calculate ρ . For this purpose, however, we need to know $T(\mathbf{R})$ and $A(\mathbf{R})$. We have therefore fitted the GIS spectra from various regions of Fig. 1a with a simple spectral model. The model consists of a plasma thermal emission²¹ of free temperature and abundance, and for a temperature of 10^8 K bremsstrahlung with a free normalization representing the emission from X-ray binaries^{4,22,23} in NGC1399. The fit was generally successful, and the plasma temperature derived is $T = (1.12\text{--}1.25) \times 10^7$ K for the whole field of view with an abundance in the range 0.2–0.4 solar.

Closer inspection of the central region using the data from the Solid State Spectrometer instrument¹⁸, which has a narrower ($22' \times 22'$) field of view but better energy resolution than the GIS, indicates a 10% temperature drop within $2.5'$ of NGC1399. The emission therefore seems to be fairly isothermal. Rosat observations yielded a more pronounced temperature decrease, by $\sim 30\%$ within ~ 35 kpc ($\sim 6'$) from NGC1399 (refs 14, 15), which may be smeared out in the ASCA data because of insufficient angular resolution.

To take into account these spectral uncertainties, here we consider three cases. Model 1 is isothermal ($T = 1.2 \times 10^7$ K), with a constant abundance ($A = 0.3$ solar), simulating the ASCA result. Model 2 assumes a constant abundance ($A = 0.3$ solar), but its temperature is given as $T(R) = [1.2 - 0.4 \exp(-R/a)] \times 10^7$ K (where $R = |\mathbf{R}|$) with $a = 15$ kpc, simulating the Rosat result¹⁴. Model 3 is an isothermal ($T = 1.2 \times 10^7$ K) model, but its abundance increases as $A(R) = 0.3 + 0.3 \exp(-R/a)$ solar towards the centre with the same a as model 2. By substituting these profiles into equations (2) and (3), we have calculated $n(\mathbf{R})$ and $\rho(\mathbf{R})$ at each grid point in three-dimensional space. As neither quantity deviates much from spherical symmetry, we have integrated $n(\mathbf{R})$ and $\rho(\mathbf{R})$ within a sphere of radius R centred on NGC1399, to

obtain the integrated plasma mass $M_{\text{gas}}(R)$ and the total gravitating mass $M_{\text{tot}}(R)$ within R . These results are shown in Fig. 2, where we have also plotted the integrated mass profile of the stellar component, $M_{\text{star}}(R)$, and that of the dark matter, $M_{\text{DM}}(R)$, obtained by subtracting M_{star} and M_{gas} from M_{tot} .

In Fig. 2, the total and dark matter distributions show a clear galaxy/cluster hierarchical structure, which is independent of which of the three spectral models is chosen, except for minor differences. As R increases,

TABLE 1 Best-fit parameters for the two β models

	Galaxy component	Cluster component
Central emissivity ($\text{erg s}^{-1} \text{pc}^{-3}$)*†	$(0.93 \pm 0.18) \times 10^{29}$	$(1.2 \pm 0.5) \times 10^{26}$
Central gas density (cm^{-3})†	$(2.3 \pm 0.2) \times 10^{-2}$	$(8.2 \pm 1.6) \times 10^{-4}$
Core radius (kpc)*	$4.8 (< 6.0)$	127 ± 6
β	0.51 ± 0.04	0.60 ± 0.04
Centre offset (kpc)‡	—	44 ± 6

Parameters are for the best-fit two-component β -model fitted to the one-dimensional brightness profile of Fig. 1b. All errors represent statistical plus systematic errors.

* The distance to the object is assumed to be 20 Mpc, yielding a plate scale of 5.8 kpc to 1 arcmin.

† The X-ray emitting gas is assumed to have a temperature of 1.2×10^7 K and an abundance of 0.3 solar.

‡ The centre of the cluster component is assumed to be offset from that of the galaxy component, in the northwest direction along the strip shown in Fig. 1a, within the sky plane.

the M_{tot} and M_{DM} profiles flatten out over $R \approx 40\text{--}50$ kpc, and then start increasing again beyond $R \approx 80$ kpc. This means that the dark matter is not only distributed on the cluster-wide scale, but also concentrated on a scale of ~ 60 kpc. This feature can be traced back to a subtle curvature change in the M_{gas} curve at $R \approx 60$ kpc, where the two β components in Fig. 1b cross over. Therefore the 'galaxy' component clearly results from an additional potential dimple around NGC1399 where an excess plasma pressure is confined; within the observational tolerances, it cannot be attributed to a central temperature decrease (model 2) or to a central metallicity enhancement (model 3).

In Fig. 2, at $R \equiv R_1 \approx 60$ kpc, there exists a well defined interface between territories of the cD galaxy and the entire cluster. For the 'galaxy' territory ($R < R_1$) we have $M_{\text{tot}} \approx 2 \times 10^{12} M_{\odot}$, $M_{\text{star}}/M_{\text{tot}} \approx 0.3$ and $M_{\text{gas}}/M_{\text{tot}} \approx 0.01$, and the 'baryon fraction' $f_B \equiv (M_{\text{star}} + M_{\text{gas}})/M_{\text{tot}} \sim 0.3$. For the 'cluster' region we have $M_{\text{tot}} \approx 2 \times 10^{13} M_{\odot}$, $M_{\text{star}}/M_{\text{tot}} \approx 0.05$, $M_{\text{gas}}/M_{\text{tot}} \approx 0.02$, and $f_B \approx 0.07$, all within $R = 300$ kpc. These values derived for $R < R_1$ and $R > R_1$ are consistent with those usually quoted for bright elliptical galaxies^{3,5} and relatively poor clusters^{2,6-8}, respectively.

Our results provide unambiguous observational evidence for the hierarchical nature of the dark matter distribution, derived assuming only hydrostatic equilibrium of the plasma. In previous studies, the double-potential structure has often been assumed^{14,24} or suggested^{8,25}, but not confirmed. The additional potential dimple associated with the central galaxy is suggested independently from measurements using gravitational lensing effects^{26,27}.

What is the implication of our results for the nature of dark matter? An orthodox interpretation may be that the distribution of dark matter fluctuates on various spatial wavelengths, which develop because of gravitational instability without strong interaction with one another. In this interpretation, a potential well due to a longer-wavelength fluctuation defines a cluster within which member galaxies form at local minima of the superposed shorter-wavelength fluctuations. A more exotic alternative, however, may be that there are two different forms of dark matter, one of which (associated with the galaxies) has been subject to some energy dissipation so that it has become more concentrated than the other (the cluster dark matter). Comparison with numerical simulations would be of great interest. $\square$

Received 3 August; accepted 12 December 1995.

- Trimble, V. A. *Rev. Astr. Astrophys.* **25**, 425–472 (1987).
- Sarazin, C. L. *X-ray Emission from Clusters of Galaxies* (Cambridge Univ. Press, Cambridge 1988).
- Forman, W., Jones, C. & Tucker, W. *Astrophys. J.* **293**, 102–119 (1985).
- Canizares, C. R., Fabbiano, G. & Trinchieri, G. *Astrophys. J.* **312**, 503–513 (1987).
- Awaki, H. *et al. Publ. astr. Soc. Japan* **46**, 165–170 (1994).
- Jones, C. & Forman, W. *Astrophys. J.* **276**, 38–55 (1984).
- Edge, A. C. & Stewart, G. C. *Mon. Not. R. astr. Soc.* **252**, 417–427 (1991).
- Thomas, P. A., Fabian, A. C. & Nulsen, P. E. *Mon. Not. R. astr. Soc.* **228**, 937–991 (1987).
- Takano, S. *et al. Nature* **340**, 289–290 (1989).
- Ferguson, H. & Sandage, A. *Astr. J.* **100**, 1–33 (1990).
- Bicknell, G. V., Carter, D., Killeen, N. E. B. & Bruce, T. E. *Astrophys. J.* **336**, 639–654 (1989).
- Killeen, N. E. & Bicknell, G. V. *Astrophys. J.* **325**, 165–179 (1988).
- Serlemitsos, P., Loewenstein, M., Mushotzky, R. F., Marshall, F. E. & Petre, R. *Astrophys. J.* **413**, 518–530 (1993).
- Rangarajan, F. V. N., Fabian, A. C., Forman, W. R. & Jones, C. *Mon. Not. R. astr. Soc.* **272**, 665–667 (1995).
- Jones, C. *et al. Astrophys. J.* (in the press).
- Nugent, J. *et al. Astrophys. J. Suppl. Ser.* **51**, 1–28 (1983).
- Ikebe, Y. *et al. Astrophys. J.* **384**, L5–L8 (1992).
- Tanaka, Y., Inoue, H. & Holt, S. S. *Publ. astr. Soc. Japan* **46**, L37–L41 (1994).
- Serlemitsos, P. J. *et al. Publ. astr. Soc. Japan* **47**, 105–114 (1995).
- Ikebe, Y. thesis, Univ. Tokyo (1995).
- Raymond, J. C. & Smith, B. W. *Astrophys. J. Suppl. Ser.* **35**, 419–439 (1977).
- Makishima, K., Ohashi, T., Hayashida, K., Inoue, H. & Koyama, K. *Publ. astr. Soc. Japan* **41**, 697–708 (1989).
- Matsushita, K. *et al. Astrophys. J.* **436**, L41–L45 (1994).
- Stewart, G. C., Fabian, A. C., Jones, C. & Forman, W. *Astrophys. J.* **285**, 1–6 (1984).
- Nulsen, P. E. J. & Böhringer, H. *Mon. Not. R. astr. Soc.* **274**, 1093–1098 (1995).
- Fort, B. & Mellier, Y. *Astr. Astrophys. Rev.* **5**, 239–292 (1994).
- Miralda-Escudé, J. & Babul, A. *Astrophys. J.* **449**, 18–27 (1994).

ACKNOWLEDGEMENTS. Development of the analysis technique was supported by the ASCA–ANL and SimASCA software development teams. We thank their members and the members of the ASCA team for spacecraft operation and data acquisition, and A. C. Fabian, Y. Suto and M. Hattori for discussions.

Rapid acceleration of the polar solar wind

R. R. Grall*, W. A. Coles*, M. T. Klingle-Smith*,
A. R. Breen†, P. J. S. Williams†, Jussi Markkanen‡
& Ruth Esser§

* Electrical and Computer Engineering, University of California, San Diego, La Jolla, California 92093-0407, USA

† Physics Department, University of Wales, Prifysgol Cymru, Aberystwyth, UK

‡ EISCAT: Geophysical Observatory, SF99600 Sodankylä, Finland

§ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

THE solar wind is a supersonic outflow of coronal plasma into interplanetary space, and is the agent that carries solar disturbances to the Earth. Direct measurements of the wind speed over a range of distances—from the orbit of Mercury¹ to beyond the outermost planets² and now over the solar poles³—show that the acceleration is largely complete by 70 solar radii ($R_{\odot}$). But there are no direct measurements nearer the Sun with which to constrain theoretical models of the acceleration. In principle, the speed of the solar wind in the acceleration region can be inferred by indirect methods such as radio scattering, but this is not straightforward as these data provide a measure of the wind properties integrated along the lines of sight. Here we report radio-scattering measurements of the speed of the south polar stream which have been corrected for this path integration, and also for the potential bias due to the presence of plasma waves. Our results indicate that the acceleration of the polar wind is almost complete by $10 R_{\odot}$, much closer to the Sun than had been expected⁴. This suggests that the acceleration of the solar wind and the heating of the solar corona occur in essentially the same region, and thus that the underlying mechanisms may be strongly linked⁵.

In a radio-scattering observation, fluctuations in the electron density act as a flow tracer⁶. They cause refractive-index fluctuations that phase-modulate the radio wave from a distant source, such as a quasar, as it passes near the Sun. Diffraction converts this phase modulation to intensity modulation as the wave continues to the Earth. We observe the drifting intensity pattern as a time series of intensity scintillation with two separated radio telescopes. The time delay between the measurements at the two antennas provides an estimate of the flow speed.

The scintillation is a random process so the time lag must be estimated using a correlation analysis. If it is characterized by a single velocity, the cross-correlation function will be a delayed replica of the autocorrelation. But the spatial intensity pattern will change as it drifts owing to turbulence and/or wave motion, so the time series observed at the two antennas will differ slightly. This will cause the peak of the cross-correlation to drop below unity and will shift it towards zero lag. Various methods of correcting for this have been discussed in the literature^{7–11}. All of these methods provide an estimate of the mean speed over the scattering region.

Observations over several solar cycles have shown that the solar wind is grossly inhomogeneous. It is now apparent that its structure is, to first order, bimodal^{1,3,12}, that is, it consists of 'fast' and 'slow' streams which are internally homogeneous. The fast wind is known to arise in regions of open magnetic field called 'coronal holes'. Slow wind is found above closed field regions such as the equatorial streamer belt which forms at solar minimum, but it is not clear where it originates. If the scattering volume includes both fast and slow wind, as it often does during the declining and minimum phases of the solar activity cycle, then the mean speed is not useful for comparison with theoretical acceleration models.

Scintillation observations made at the three-station EISCAT¹¹ observatory in northern Scandinavia began to show unusual

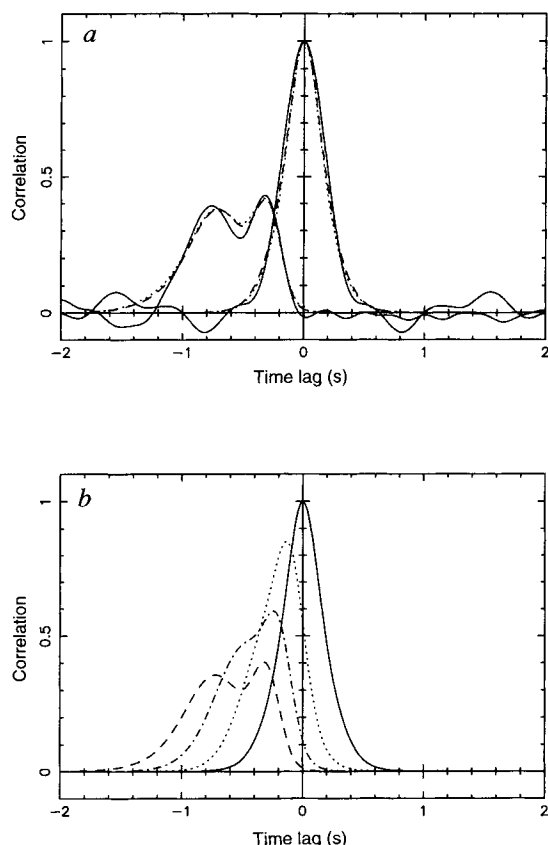


FIG. 1 *a*, Intensity correlation functions from EISCAT on 28 April 1994. The source was 0323+055 at -52° latitude, and the baseline was 240 km radial and 3 km tangential. The observations are shown by solid lines, two model fits are plotted as dotted (model 1) and dashed (model 2) lines. The fast wind occupies a sector of the path subtending from -45° to $+90^\circ$ (with respect to the perpendicular) at the Sun. *b*, The effect of baseline length on a solar wind model (1) which fits the observations shown in *a*. The autocorrelation is a solid line; the cross-correlations have been calculated with the actual baseline of 240 km (dashed line), and also with 160-km (dot-dashed) and 80-km (dotted) baselines. Model parameters are: (1) axial ratio = 1.35, mean speed in the slow wind $V_{\text{slow}} = 500 \text{ km s}^{-1}$, transverse random velocity in the slow wind $V_{\text{perp-slow}} = 50 \text{ km s}^{-1}$, mean speed in the fast wind $V_{\text{fast}} = 848 \text{ km s}^{-1}$, transverse random velocity in the fast wind $V_{\text{perp-fast}} = 34 \text{ km s}^{-1}$, ratio of δN_e^2 in the slow and fast wind $\delta N_e^2 \text{ slow/fast} = 23$; (2) axial ratio = 1.35, $V_{\text{slow}} = 500 \text{ km s}^{-1}$, $V_{\text{perp-slow}} = 20 \text{ km s}^{-1}$, $V_{\text{fast}} = 777 \text{ km s}^{-1}$, $V_{\text{perp-fast}} = 136 \text{ km s}^{-1}$, $\delta N_e^2 \text{ slow/fast} = 14$.

cross-correlations with two well separated peaks during the spring of 1994. An example is shown in Fig. 1*a*. We assume that this is caused by the combined effects of fast and slow wind. To confirm this we have examined the underlying corona to see if the path crosses regions where one could expect both fast and slow wind. This comparison is shown in Fig. 2. Indeed this path is centred on a large polar coronal hole and it continues into an equatorial streamer.

We have estimated both components using a simple model with a spherically diverging geometry. The fast and slow wind are characterized by a mean speed and transverse random velocity which are independent of distance, and a density variance δN_e^2 which drops with distance as $\delta N_e^2 \propto R^{-4}$. Only the ratio of the density variance in fast and slow wind enters the calculation, so there are five free parameters. The two models that provide upper and lower bounds on the fast speed are plotted with the measured correlations in Fig. 1*a*. The location of the fast-slow boundary was taken from the coronal measurements shown in Fig. 2.

Double-peaked cross-correlations have not been seen before because the inter-antenna baselines available at EISCAT (up to 400 km) are much longer than have been used before. This effect is demonstrated in Fig. 1*b* where we have plotted a model fit to the data of Fig. 1*a*, both with the actual baseline (240 km) and as it would appear with two shorter baselines (160 and 80 km). We see that shorter baselines would not resolve the two peaks, and that only a weighted average of the fast and slow speeds could be estimated.

When the intensity variance is small, the correlation functions can be calculated using the Born approximation¹³. In this case the observed correlations are simply a weighted average of contributions from all along the path. However, if the scattering is strong, the path integral appears in an exponent¹⁴ and double peaks cannot occur. Thus the fast and slow contributions can only be separated when the scintillation is weak. At the EISCAT frequency of 933 MHz, the solar distance must exceed $30 R_\odot$ in the equatorial region, and $15 R_\odot$ in the polar streams (in which the electron density fluctuations are much weaker¹⁵). However, the antennas of the Very Long Baseline Array (VLBA) can observe at much higher frequencies¹⁶. The VLBA observations presented here were made at 8 GHz but it should be possible to work as close as $2 R_\odot$ at 15 GHz.

A weakness in scintillation velocity measurements is that they measure the speed of the density fluctuations not the flow velocity itself. If, for example, the density fluctuations were caused by outward propagating sound waves, the apparent speed would be the flow speed plus the sound speed. There is no reason to believe that the density fluctuations in the solar wind are caused by sound waves, but it is known that Alfvén waves are prevalent¹⁷, particularly in the fast polar streams¹⁸. These are magnetic field shear waves and do not drive density fluctuations directly. But large-scale Alfvén waves propagating in a medium containing isotropic

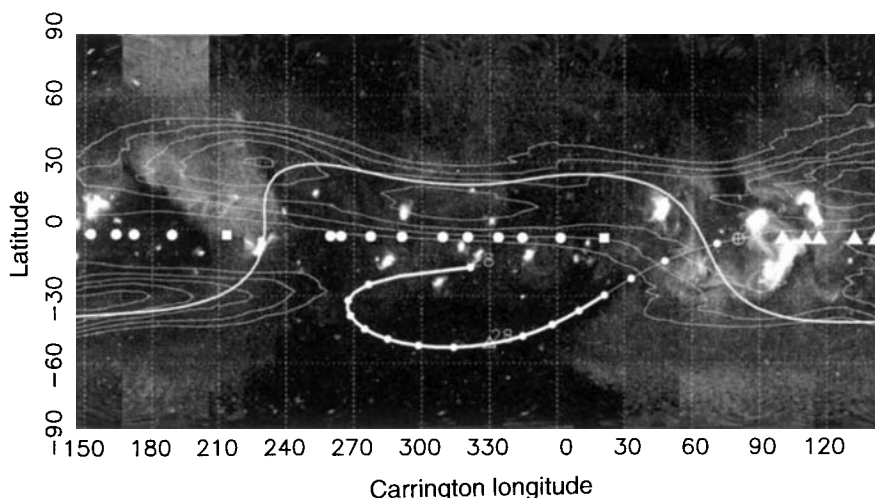
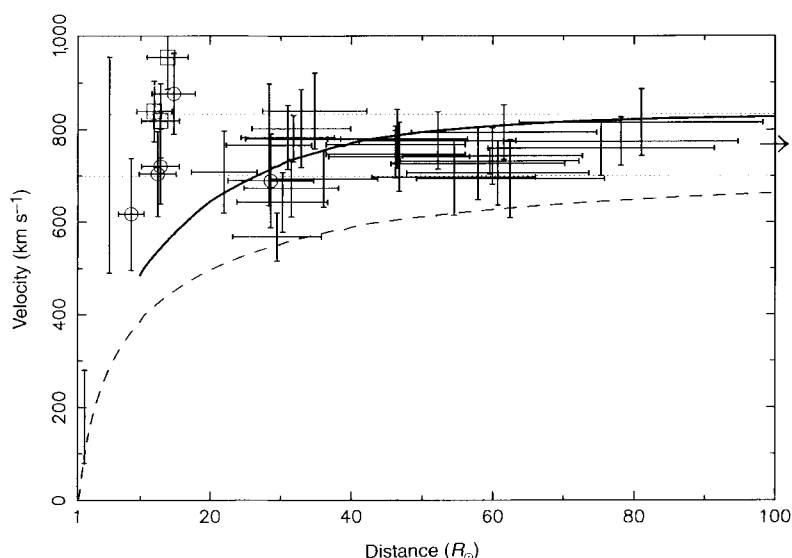


FIG. 2 Comparison of the radio path with coronal measurements in heliographic (Carrington) coordinates. The Yohkoh soft X-ray emission is plotted in a grey-scale. The density estimates made from the Mauna Loa data are plotted as contours. The magnetic neutral line estimated by the Wilcox Solar Observatory group at Stanford University is plotted as a heavy solid line. The IMP8 slow speeds are marked as large triangles ($< 500 \text{ km s}^{-1}$), the fast speeds as large circles ($> 600 \text{ km s}^{-1}$), and the intermediate speeds as large squares. The path is marked with an open triangle at the point of closest approach and with small circles at 10° intervals. The Earth end is marked with a $\oplus$. The region that we assume is fast is a heavy line and the slow region a light line.

FIG. 3 Measurements of the apparent flow speed versus distance. The vertical bar on each data point is the 90% confidence limit, and the horizontal bar indicates the distance range over which the scintillation estimate is averaged. This distance range is from 1.0 to 1.54 times the distance of closest approach. The EISCAT measurements in the south polar stream between April and September 1994 are marked with crossed bars. The VLBA measurements are marked with circles (from the south pole) or squares (from the north pole). The south VLBA observations at 6.8 and 22.5 $R_{\odot}$ are from 15 May 1994, the others are from 20–22 October 1995. The north VLBA observation at 9.4 $R_{\odot}$ is from 23 April 1994, the others are from 1–2 August 1995. The estimates²³ derived from SPARTAN 201–01²² on 11–12 April 1993 are plotted as vertical bars at 2 and 5.5 $R_{\odot}$. The upper and lower bounds of the Ulysses measurements south of -60° latitude (17 April to 1 December 1994) are plotted as horizontal dotted lines. At this time Ulysses was much further out; between 375 and 700 $R_{\odot}$. The mean Ulysses speed is marked with an arrow at 100 $R_{\odot}$. The flow speed of a wave-assisted acceleration model⁴ is plotted as a dashed line, and the apparent scintillation velocity calculated from this model, including the wave-bias, is plotted as a heavy solid line.



small-scale density fluctuations will cause an apparent transverse random velocity, which is just the wave amplitude δV_{pert} .

If the density fluctuations are field-aligned, the apparent scintillation speed depends on the amplitude and speed of the Alfvén waves, and the axial ratio of the density fluctuations¹⁹. In this case the correlation functions alone do not provide enough information to estimate both the amplitude and speed of the waves. We can calculate the apparent velocity from a theoretical model and compare the measured apparent velocity with the theoretical prediction, but we cannot estimate the actual flow speed independently of the theoretical model. For the observations presented here the maximum bias is $\sim 15\%$ of the flow speed, but this bias may become more serious for observations inside 6 $R_{\odot}$, where the anisotropy is much larger²⁰.

The estimation error in our fitting procedure depends more on systematic error than statistical error. The systematic error is due to a strong correlation between the effects of three model parameters: the two random velocities, and the density variance ratio. All three affect the height of the cross-correlation peaks, but observations that are entirely in the slow wind show random velocities between 20 and 100 km s^{-1} . Using these limits we can tighten the confidence limits on the remaining parameters. We cannot place similar limits on the random velocity in the fast wind, because we have few observations completely dominated by the fast wind.

The observations presented here are from the polar regions where the centre of the path is in the fast wind and the ends are often in slow wind. In most cases the fast wind is more dominant than it is in Fig. 1. For polar observations the exact location of the fast–slow boundary does not affect the fast peak in the cross-correlation. It does affect the slow peak, because the apparent speed of the slow contribution is reduced by a ‘cosine’ projection effect. In such cases the slow wind is poorly determined, but the fast wind is well determined because the correction for the slow wind is smaller.

Estimates of the south polar wind speed are plotted in Fig. 3 as crossed bars. The vertical error bar is dominated by the systematic error discussed earlier. We have included VLBA measurements in the north polar stream because of the paucity of observations near the Sun, the obvious importance of the innermost points, and the fact that recent observations by the Ulysses spacecraft have shown that the speed distribution over the north pole is indistinguishable from that under the south pole²¹. The maximum, minimum and mean speeds from Ulysses south of -60° are also plotted in Fig. 3.

Estimates of the flow speed near the Sun can be derived from density measurements under the assumptions that mass flux is conserved and the expansion geometry is known. Density estimates made from the white light coronagraph on the Spartan 201–01 space shuttle mission²² were used by Habbal *et al.*²³ to obtain the

speed estimates at 2 $R_{\odot}$ and 5.5 $R_{\odot}$ plotted in Fig. 3. At 5.5 $R_{\odot}$ the flow was assumed radial, whereas at 2 $R_{\odot}$ both radial and diverging Munroe and Jackson²⁴ geometries were used.

A two-fluid acceleration model which includes Alfvén waves⁴ is plotted in Fig. 3 as a dashed line. In this model, the waves propagate according to the WKB approximation and are not damped. The model was not optimized to provide the best-fit, but provides a useful comparison and allows us to estimate the maximum bias caused by Alfvén waves. The apparent scintillation speed was calculated from the theoretical model and is plotted as a heavy solid line. If the model is correct this heavy line should run through the observations. The difference between the dashed line and the heavy line is the bias due to Alfvén waves propagating through field-aligned density irregularities.

The wave-driven acceleration model matches the Ulysses observations at large distances, but does not match the observations near the Sun, even with the maximum correction for wave-bias. It seems unlikely that models of this class which use the pressure gradient force of undamped Alfvén waves to help accelerate the wind, will be able to duplicate this rapid acceleration in the region below 10 $R_{\odot}$. It also suggests that the mechanisms by which the corona is heated and the solar wind accelerated are not as independent as had been suggested²⁵. It is clear that further indirect observations in the range 2–4 $R_{\odot}$ will be very valuable. It is also clear, in view of the potential bias in the radio observations, that eventually direct observations will be needed inside 6 $R_{\odot}$. $\square$

Received 10 July; accepted 6 December 1995.

- Schwenn, R. in *Physics of the Inner Heliosphere* (eds Schwenn, R. & Marsch, E.) 99–181 (Springer, Heidelberg, 1990).
- Richardson, J. D., Paularena, K. I., Lazarus, A. J. & Belcher, J. W. *Geophys. Res. Lett.* **22**, 325–328 (1995).
- Phillips, J. L. *et al.* *Geophys. Res. Lett.* **21**, 1105–1108 (1994).
- Esser, R., Leer, E., Habbal, S. R. & Withbroe, G. L. *J. geophys. Res.* **91**, 2950–2960 (1986).
- Hollweg, J. V. *J. geophys. Res.* **91**, 4111–4125 (1986).
- Dennison, P. A. & Hewish, A. *Nature* **213**, 343–346 (1967).
- Briggs, B. H., Phillips, G. J. & Shinn, D. H., *Proc. R. Soc. Lond. B* **63**, 106–121 (1950).
- Little, L. T. & Ekers, R. D. *Astr. Astrophys.* **10**, 306–309 (1971).
- Armstrong, J. W. & Coles, W. A. *J. geophys. Res.* **77**, 4062–4610 (1972).
- Coles, W. A. & Kaufman, J. J. *Radio Sci.* **13**, 591–597 (1978).
- Bourgeois, G. *et al.* *Astr. Astrophys.* **144**, 452–462 (1985).
- Axford, W. I. & McKenzie, J. F. in *Proc. of Solar Wind Eight* (eds Winterhalter, D., Gosling, J., Habbal, S. W., Kurth, W. & Neugebauer, M.) (AIP, New York, in the press).
- Lee, R. W. & Harp, J. C. *Proc. I. E. E.* **57**, 375–394 (1969).
- Prokhorov, A. M., Bunkin, F. V., Gochelashvili, K. S. & Shishov, V. I. *Proc. I. E. E.* **63**, 790–811 (1975).
- Coles, W. A., Grall, R. R., Klinglesmith, M. T. & Bourgeois, G. *J. geophys. Res.* **100**, 17069–17079 (1995).
- Kellermann, K. I. & Thompson, A. R. *Science* **229**, 123–130 (1985).
- Belcher, J. W. & Davis, L. J. *Geophys. Res.* **76**, 3534–3563 (1971).
- Smith, E. J. *et al.* *Space Sci. Rev.* **72**, 165–170 (1995).
- Coles, W. A. in *Wave Propagation in Random Media (Scintillation)* (eds Tatarski, V. I., Ishmaru, A. & Zavorotny, V. U.) 156–168 (SPIE, Bellingham, WA, 1993).
- Armstrong, J. W., Coles, W. A., Kojima, M. & Rickett, B. J. *Astrophys. J.* **358**, 685–692 (1990).

21. Phillips, J. L. et al. in *Proc. of Solar Wind Eight* (eds Winterhalter, D., Gosling, J., Habbal, S. R., Kurth, W. & Neugebauer, M.) (AIP, New York, in the press).
22. Fisher, R. R. & Guhathakurta, M. *Space. Sci. Rev.* **70**, 267–272 (1994).
23. Habbal, S. R., Esser, R., Guhathakurta, M. & Fisher, R. R. *Geophys. Res. Lett.* **22**, 1465–1468 (1995).
24. Munroe, R. H. & Jackson, B. V. *Astrophys. J.* **213**, 874–886 (1977).
25. Esser, R. & Habbal, S. in *Proc. of Solar Wind Eight* (eds Winterhalter, D., Gosling, J., Habbal, S. R., Kurth, W. & Neugebauer, M.) (AIP, New York, in the press).

ACKNOWLEDGEMENTS. We thank our colleagues at EISCAT and the VLBA who helped make the observations. We are also grateful to colleagues who provided that the Yohkoh images, the Mauna Loa data, the WSO neutral line estimates, and the Ulysses velocities. Discussions with B. Rickett were very helpful in clarifying the effect of Alfvén waves. This work was partially supported by the US NSF.

Viscoelastic effects in the spreading of liquids

Alain Carré*, Jean-Claude Gastel*
& Martin E. R. Shanahan†

* Centre de Recherche Corning Europe, 7 bis, Avenue de Valvins, 77210 Avon, France

† Centre National de la Recherche Scientifique, Ecole Nationale Supérieure des Mines de Paris, Centre des Matériaux P.M. Fourn, BP 87, 91003 Evry Cedex, France

THE spreading of liquids on solid substrates is important in many natural and industrial processes, including eye irrigation, adhesive bonding, tertiary oil recovery and the application of insecticides, paints and inks. A liquid drop placed on a flat, smooth, horizontal solid spreads as excess capillary potential energy is dissipated during the motion. If the solid is rigid, energy is dissipated entirely by viscous flow within the liquid¹. For a non-rigid solid, however, a local microscopic deformation, or 'wetting ridge'^{2,3}, forms near the wetting front, and its motion also induces viscoelastic dissipation^{4–6}. Here we report direct evidence of the wetting ridge obtained using scanning interferometric microscopy. Our experiments demonstrate, for liquids and solids with properties representative of practically important substances, that the mesoscopic surface deformation can significantly affect macroscopic phenomena. For instance, the formation of a nano-scale wetting ridge can increase the spreading time of a drop by considerably more than an order of magnitude, with potentially important consequences in technological and natural settings.

Under classic conditions of wetting on a smooth, horizontal and hard solid, the kinetics of spreading is controlled essentially by a dynamic energy balance between the rate of restitution of capillary potential energy and viscous dissipation occurring through shear motion within the liquid¹. Thus, if a small drop is placed on a solid surface, the contact angle, $\theta(t)$, at time t , before equilibrium is attained, is greater than the static value, θ_0 , which satisfies Young's equation $\gamma_{SV} = \gamma_{SL} + \gamma_{LV} \cos \theta_0$, where γ_{ij} are the interfacial free energies between solid (S), liquid (L) and vapour (V) phases. This capillary imbalance leads to a spreading force of $\gamma_{LV}[\cos \theta_0 - \cos \theta(t)]$ and as the wetting front advances, work is expended and dissipated by viscous shear within the liquid.

When liquids are put into contact with soft solids such as rubbers, elastomers, gels or living tissues, local deformation may result near the SLV triple line, because of the component of liquid surface tension, $\gamma_{LV} = \gamma$, acting perpendicularly to the undisturbed substrate, $\gamma \sin \theta_0$ (refs 2,3; Fig. 1). The elastic displacement of the solid, or wetting ridge, has a height, h_0 , of the order $\gamma \sin \theta_0 / G$, where G is the shear modulus of the solid^{3,4}.

Direct evidence for the existence of the wetting ridge is illustrated in Fig. 2. We investigated the topographical modification of a silicone elastomer substrate with a low shear modulus, G , near the triple line of a liquid drop, using scanning white-light interferometric microscopy. The height of the wetting ridge, h_0 , seems to be about 30 nm, in good agreement with the calculated value ($h_0 \approx \gamma \sin \theta_0 / G$) of 35 nm. Moreover, the theoretical profile

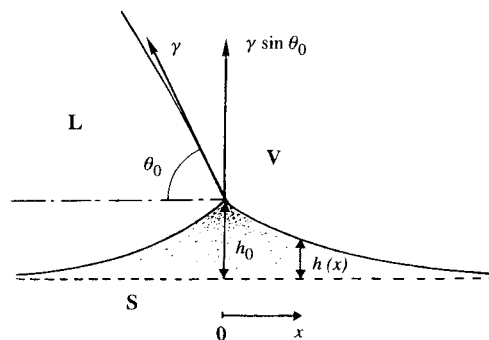


FIG. 1 Schematic representation of the wetting ridge near the triple line solid/liquid/vapour (SLV). The dotted line represents the solid without deformation caused by the vertical component of the liquid surface tension, $\gamma \sin \theta_0$, θ_0 being the contact angle at equilibrium; h_0 and $h(x)$ define the vertical displacements of the solid surface under the action of $\gamma \sin \theta_0$.

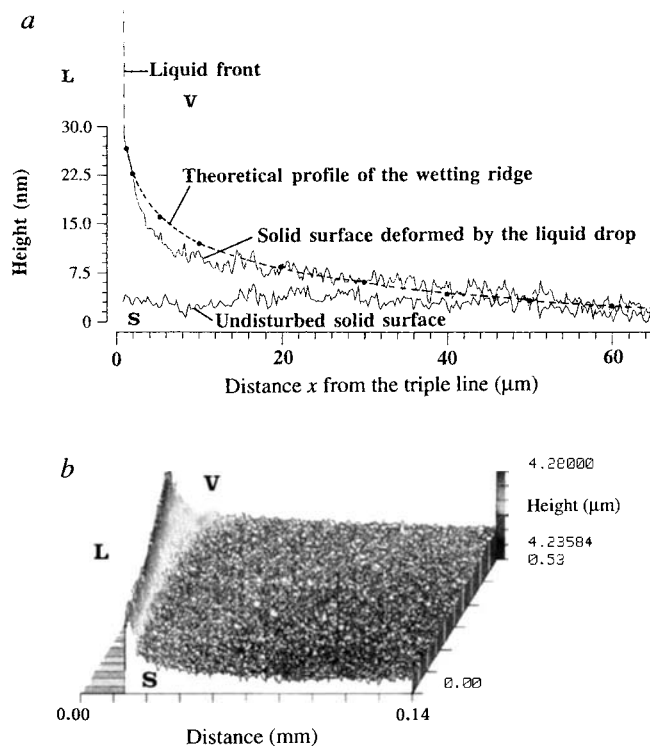


FIG. 2 Topographical modification of a soft elastomer surface (rather like a rubber band; silicone RTV 615, General Electric, $G = 0.63$ MPa, glass transition temperature $T_g = -120^\circ\text{C}$) near the triple line of a liquid drop (tricresylphosphate, Aldrich Chemical Company Inc.), measured with a scanning white-light interferometric microscope (Newview 100, Zygo) at 30°C . The principle of the instrument is derived from a classical interferometric measurement system. A light beam interferes with a reference beam. The image of the resulting interferogram is collected by a low-noise camera, and after processing, data are stored in a computer memory. To cover the domain of surface topographic variations, the interferometric objective is moved along its axis by means of a piezoelectric translation stage. Finally, the set of interferograms acquired and stored in the memory is processed using a powerful algorithm to construct a three-dimensional image of the surface. The use of a white-light source together with this algorithm permits the exploration of up to $100\text{-}\mu\text{m}$ vertical topographic variation with a 0.1-nm resolution. For this analysis, the liquid was dyed with 0.4 wt% of solvent green-3 (Aldrich Chemical Company Inc.) to obtain the required opacity. The arithmetic mean roughness of the silicone surface is about 2 nm. The liquid contact angle at equilibrium is 51° and its surface tension is 28.5 mN m^{-1} . a, Dotted line represents the theoretical profile of the wetting ridge defined⁴ by $h(x) \approx h_0 \ln(d/x)/(2\pi)$, with $h_0 \approx 35$ nm and $d \approx 60$ μm . b, Three-dimensional picture (oblique plot) of the wetting ridge obtained directly from the interferogram.

of the wetting ridge defined by the vertical displacement, $h(x)$, where x is the distance from the triple line, fits well with the experimental topographical level (Fig. 2). It can be shown that, for an elastomer,

$$h(x) \approx \frac{\gamma \sin \theta_0}{2\pi G} \ln\left(\frac{d}{x}\right), \quad x > \varepsilon \quad (1)$$

where d represents the distance along the line of action of γ at which it may be assumed that no displacement occurs within the solid⁴, and ε is a small cutoff distance near the triple line below which the behaviour of the solid is no longer linearly elastic (ε is a few nanometres)³. Therefore the displacement, $h(x)$, may be estimated using the expression $h(x) \approx h_0[\ln(d/x)]/(2\pi)$; we estimated d to be about 60 μm .

The observation of the wetting ridge and of its predicted profile and magnitude justify the following theoretical description of its involvement in the dynamics of wetting on soft solids.

Before equilibrium is attained, when the SLV triple line is moving at speed U , corresponding to a drop spreading with contact angle $\theta(t)$ decreasing towards θ_0 (Figs 3a and 4a), the wetting ridge moves with the triple line, and work $\dot{E} = \gamma^2 U / 2\pi G \varepsilon$ is done per unit time and per unit length of triple line. But with viscoelastic solids such as rubbers, a fraction Δ of the elastic energy is dissipated. The fraction Δ is rate-dependent, as was shown some years ago in adhesion experiments⁷⁻⁹, and takes the form

$\Delta = (U/U_0)^n$, where U_0 and n are constants that define the damping properties of the solid^{5,6}.

The rate at which capillary free energy is released per unit length of triple line (the 'motor') is given by $\dot{F} = \gamma U [\cos \theta_0 - \cos \theta(t)]$, this term being the product of spreading force and spreading speed. This work is consumed, totally for hard solids and partially for rubbers, by viscous dissipation¹, $T\dot{S} = 3\eta l U^2 / \theta(t)$, where $T\dot{S}$ is the viscous dissipation, η is liquid viscosity and l is a logarithmic factor, approximately constant, involving cutoff distances to the dissipation zone of the drop.

Considering the global energy balance involving both viscous and viscoelastic dissipation phenomena, we obtain

$$\cos \theta_0 - \cos \theta(t) = \frac{3\eta l U}{\gamma \theta(t)} + \frac{\gamma}{2\pi G \varepsilon} \left(\frac{U}{U_0}\right)^n \quad (2)$$

which is the basic equation governing liquid spreading in the general case. Typically, $n \ll 1$, and thus at sufficiently high speeds, the first expression on the right-hand side of equation (2) dominates, whereas at lower speeds the second expression governs behaviour. The latter situation corresponds to the experimental conditions considered here, so the viscous dissipation term may be neglected. Viscoelastic dissipation is the primary 'brake' controlling spreading speed (that viscosity becomes unimportant has been shown using liquids of very different viscosity⁶).

In Figs 3 and 4, results are given that demonstrate the very

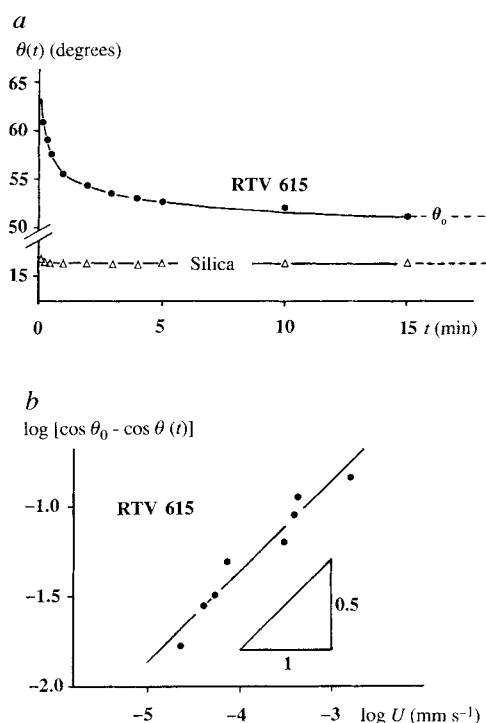


FIG. 3 a, Evolution of the contact angle $\theta(t)$ of tricesylphosphate (drop volume 2 μl) as a function of time after deposition, t , on a flat, smooth, horizontal surface of a silicone elastomer (RTV 615) at 30 °C. The contact angle at equilibrium, θ_0 (51°), is attained after about 15 minutes (compared with about 20 seconds on silica, the rigid solid). This 'abnormal' behaviour is attributed to local deformation of the elastomer near the wetting front—the 'wetting ridge'—caused by the vertical component of liquid surface tension, γ , and slowing down the spreading process. To avoid any artefact produced by a (slight) degree of swelling of the elastomer by tricesylphosphate, the substrate was immersed in liquid several days before the wetting experiment until equilibrium was noted by gravimetry (in all cases, equilibrium swelling was less than 1% by weight). The solid was then carefully dried before depositing drops of the same liquid to do the spreading experiment (after contact with the elastomer $\gamma = 28.5 \text{ mN m}^{-1}$ and $\eta = 65 \text{ cP}$ at 30 °C; the liquid behaves in flow rather like olive oil). b, Linear variation of $\log[\cos \theta_0 - \cos \theta(t)]$ as a function of the logarithm of liquid spreading speed, U . This corroborates the existence of viscoelastic braking for

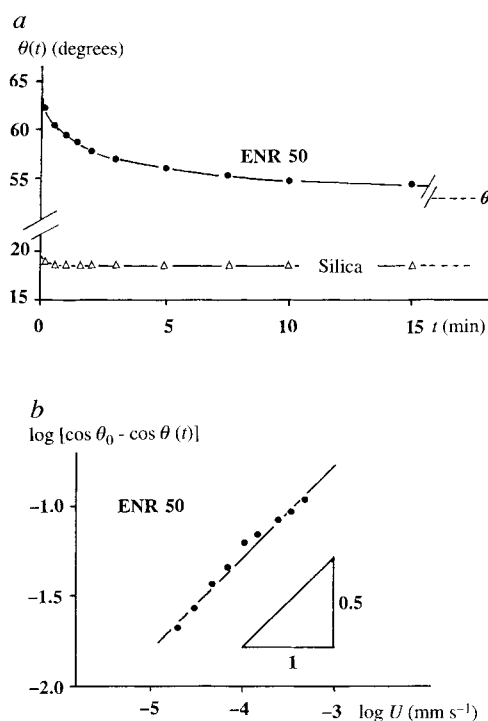


FIG. 4 a, Evolution of the contact angle $\theta(t)$ of a 2- μl drop of formamide as a function of time after deposition, t , on a flat, smooth, horizontal surface of epoxidized natural rubber (ENR 50, Malaysian Rubber Producers' Research Association, $G = 0.37 \text{ MPa}$, $T_g = -20^\circ\text{C}$) at 20 °C. The contact angle at equilibrium, θ_0 (53°), is attained after 1 hour (compared with about 20 seconds on silica). The interpretation is as for Fig. 3. In this case, after contact with the elastomer $\gamma = 45.9 \text{ mN m}^{-1}$ and $\eta = 3.2 \text{ cP}$ at 20 °C; the liquid behaves in flow rather like water. b, Linear variation of $\log[\cos \theta_0 - \cos \theta(t)]$ as a function of the logarithm of the liquid spreading speed, U . This corroborates the phenomenon of viscoelastic braking for formamide spreading on ENR 50 rubber.

tricesylphosphate spreading on silicone rubber. Spreading rate was calculated directly from observations of drop contact radius, r , as a function of time, t . The gradient $n = 0.50$ is in good agreement with other dynamic wetting and adhesion experiments^{5-7,9}. This shows that a mesoscopic surface strain leads to a macroscopic modification of wetting behaviour.

different behaviour observed during spreading on two elastomeric solids compared with that on a hard solid, silica. Figure 3a shows the variation of contact angle of tricresylphosphate (a nonvolatile and medium-viscosity liquid) as a function of time after deposition, t , on the flat, smooth, horizontal surface of a soft silicone elastomer. Figure 3b illustrates the relationship between $\log[\cos \theta_0 - \cos \theta(t)]$ and $\log U$ corresponding to equation (2) (ignoring the viscosity component) and leading to a gradient n of 0.50. Linearity, with a correlation coefficient of the least-squares linear regression, r , close to 1, and the value of $n = 0.50$ are in accordance with viscoelastic dissipation^{5,6,9} in the wetting ridge. Figure 4 shows similar behaviour for the spreading of formamide on an epoxidized natural rubber. This liquid has a viscosity close to that of water, indicating that viscosity is not the governing factor. Moreover, it can be seen from equation (2) that if viscosity were the overriding 'brake', the gradient of the lines shown in Figs 3b and 4b would be (approximately) unity. This clearly not so.

The results demonstrate that the spreading of a liquid on an elastomer is largely dependent on viscoelastic dissipation in the wetting ridge of the substrate near the triple line. Under 'classic' conditions, spreading behaviour is assumed to depend entirely on liquid properties. It is now clear that the spreading kinetics of a liquid on a solid can be controlled by the mechanical properties of the substrate itself under some circumstances. The theoretical model recently proposed to interpret wetting dynamics on soft

solids^{5,6} is now supported by the direct observation of the wetting ridge near to the solid/liquid/vapour triple line.

We have shown the wetting ridge to be important with other soft solids and liquids including water; its existence may therefore have far-reaching consequences, especially in natural systems. For example, human skin, arteries and muscle have an initial elastic modulus of about 100 kPa (ref. 10) and thus with water a wetting ridge about 1 μm high may be expected. This, combined with the known viscoelastic nature of soft human tissues¹⁰, shows that dissipative spreading may be important in a biological context. It is probable that viscoelastic 'braking' occurs in certain biological processes such as those involving membranes. There are many other circumstances involving soft solids, where the observed phenomena may also be pertinent (for instance in printing, painting, adhesive bonding, contact lenses and so forth). $\square$

Received 19 May; accepted 19 December 1995.

1. de Gennes, P. G. *Rev. mod. Phys.* **57**, 827 (1985).
2. Lester, G. R. *J. Coll. Sci.* **16**, 315 (1961).
3. Shanahan, M. E. R. & de Gennes, P. G. *Adhesion* **11** (ed. Allen, K. W.) 71 (Elsevier Applied Science, London, 1987).
4. Shanahan, M. E. R. *J. Phys. D: appl. Phys.* **21**, 981 (1988).
5. Shanahan, M. E. R. & Carré, A. *Langmuir* **10**, 1647 (1994).
6. Carré, A. & Shanahan, M. E. R. *Langmuir* **11**, 24 (1995).
7. Gent, A. N. & Schultz, J. *J. Adhesion* **3**, 281 (1972).
8. Andrews, E. H. & Kinloch, A. J. *Proc. R. Soc. London* **A332**, 385 (1973).
9. Maugis, D. & Barquins, M. *J. Phys. D: appl. Phys.* **11**, 1989 (1978).
10. Duck, F. A. *Physical Properties of Tissue* (Academic, London, 1990).

Large-scale mapping of ocean surface currents with dual over-the-horizon radars

T. M. Georges*, J. A. Harlan† & R. A. Lematta‡

* Environmental Technology Laboratory, National Oceanic and Atmospheric Administration, Boulder, Colorado 80303, USA

† Cooperative Institute for Research in Environmental Sciences, NOAA/University of Colorado, Boulder, Colorado 80303, USA

‡ Research Development Test and Evaluation Division, Naval Command Control and Ocean Surveillance Center, San Diego, California 92152, USA

DETAILED information about near-surface ocean currents is needed for effective fisheries management, pollution mitigation, search and rescue, and climate studies, but the present generation of measurement techniques provides only limited spatial and temporal resolution or coverage^{1,2}. In near-coastal environments, pairs of shore-based high-frequency radars have been used to map surface currents over an area of a few hundred square kilometres^{3,4}. The potential for mapping open-ocean current fields has been demonstrated using military high-frequency radars that can be used to 'see' over the horizon for thousands of kilometres by reflecting signals off the ionosphere. But using one radar, only one current component can be mapped by this method⁵. Here we report the mapping of surface-current vectors obtained from simultaneously employing two such radar systems with overlapping coverage. We obtain a current map in the Florida Straits, about 1,500 km from the radars, covering two 70,000 km² areas at a resolution of 10 km and 0.1 m s⁻¹. As it employs only about 2% of the radars' potential coverage, the test shows the potential of this technique for mapping the more energetic features of ocean circulation—such as boundary currents and mesoscale eddy systems—over vast ocean areas.

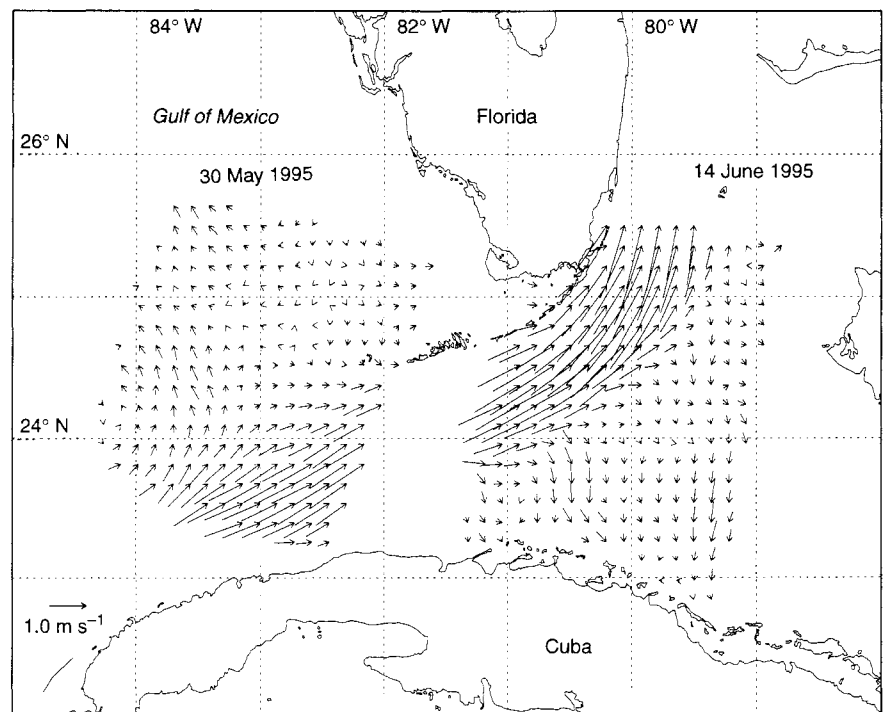
Decametric radio waves interact with the ocean primarily by Bragg-resonant scattering from surface gravity waves. For the special case of radio waves scattered back towards their source, the Bragg-resonant ocean waves are aligned perpendicular to the

line-of-sight direction and have a wavelength equal to half the radio wavelength. For example, ten-metre wavelength ocean waves backscatter 15-MHz radio waves. Radially travelling gravity waves Doppler-shift the backscattered energy by an amount proportional to their phase velocity. The spectrum of a sea echo thus consists of two sharp Doppler-shifted lines: a positive-shifted line from the Bragg-resonant waves travelling towards the radio source and a negative-shifted line from those receding from the source. If, in addition, surface currents transport the gravity waves, an extra Doppler shift is imposed that is proportional to the radial component of the current. These principles form the basis of HF (high frequency, that is, 3–30 MHz) current-mapping radars, which are now commercially available^{3,4}. A single radar maps the radial current component; two radars are required to map surface-current vectors. Relying on line-of-sight and surface-wave illumination of the sea surface, a pair of such shore-based radars can effectively map surface currents, for example in bays and estuaries, up to a range of about 70 km.

Because HF radio waves reflected from the ionosphere can reach great ranges (up to 3,500 km in one 'hop'), it was at first thought that the range of current-mapping radars could be usefully extended in this manner to cover millions instead of hundreds of square kilometres of open ocean. Two obstacles stand in the way, however. The first is the need for very large aperture antenna arrays. Beams of 1° or less in azimuth are required to resolve the most interesting ocean current features at ranges of 1,000 km or more. For a frequency of 15 MHz, for example, a 1° beam would require a kilometre-long array. The second obstacle is the distortion that the sea echo suffers after two reflections from the ionosphere (one outgoing and the other on the return path). Ionospheric reflections spread and shift the sea-echo spectrum, making it difficult to measure exact Doppler shifts or to distinguish Doppler shifts caused by the currents from those caused by ionospheric motions.

Large-aperture HF over-the-horizon radars developed and deployed as part of military air-defence systems satisfy the requirement for narrow radar beams. Since 1989, the US Air Force and Navy have permitted the National Oceanic and Atmospheric Administration to conduct limited ocean-monitoring tests with their over-the-horizon (OTH) radar systems, some of which have been deactivated with the end of the Cold War. The details of

FIG. 1 Results of the first attempt to map ocean surface currents using two over-the-horizon (OTH) radars. The two US Navy ROTHr radars, one in Virginia and the other in Texas, reflected decametric radio waves off the ionosphere to illuminate these two 70,000-km² ocean regions in the Straits of Florida. The two regions were mapped 15 days apart. Each radar mapped radial current components, and the two measurements were combined to form current vectors. The arrows show the direction of surface flow, and their lengths are proportional to the magnitude of the surface current (averaged over 2-m depth) at the arrow tip. A maximum current of 2.0 m s⁻¹ is present just off the east coast of Florida. The areas illuminated for this test are ~2% of the ocean area covered by the radars, which includes the entire Caribbean Sea and the southern Gulf of Mexico.



these radars, as they pertain to ocean monitoring, are unclassified⁶.

We deal with the second obstacle, ionospheric distortion, by recognizing that the space and time scales of ionospheric variability are generally different from those of ocean currents. Much of the Doppler 'noise' superimposed on the sea echoes is caused by ionospheric motions with timescales of minutes to hours, whereas most open-ocean currents vary on timescales of days or more. This allows us to use spatial and temporal filters to separate the two effects. For example, current features that persist from day to day are presumed not to be of ionospheric origin. Changes in the ionosphere on a diurnal timescale are roughly predictable and impose biases that are nearly uniform over large ocean areas. Such biases can be removed by high-pass spatial filtering, in which a time-dependent area-average current is subtracted from each measurement. In addition, we take advantage, whenever possible, of the stable reflections provided by the daytime E and F1 layers of the ionosphere, as well as zero-Doppler references provided by nearby land echoes. When such stable echoes are not available, current measurements may have to be averaged over several days to obtain suitable accuracy.

Soon after the US Navy deployed the second of its two relocatable over-the-horizon radar (ROTHR) systems in Texas, we set up both radars to illuminate the same patch of ocean in the Florida Straits, about 1,500 km from the Texas and Virginia radars. A frequency of 14.5 MHz was selected using real-time sweep-frequency diagnostic soundings over both oblique radar paths. A 24.6-s Doppler frequency transform was used, giving a nominal radial-current resolution of about 0.42 m s⁻¹. Using parabolic interpolation to locate the spectral peaks more accurately, we can achieve a current-speed resolution exceeding 0.1 m s⁻¹. Greater resolution is possible using longer coherent integration, but the limit imposed by ionospheric shifting and smearing is not yet known.

The ROTHR achieves its nominal 0.5° angular resolution with 2.58-km-long linear phased receiving arrays consisting of 372 twin-monopole elements. Range resolution is achieved by transmitting a continuous 25-kHz frequency-modulated waveform. A radar resolution cell on the ocean surface is therefore about 6 km in range by about 15 km in azimuth, for the frequency and range used in this test. Only sea echoes that pass a built-in quality test are used for current measurements. The quality index is derived from the

sharpness of the sea-echo spectrum, which can be degraded by different ionospheric motions over multiple radar paths, or by interference from other users of the HF radio spectrum.

The first test used E-layer propagation near midday on 30 May 1995, and data collection from both radars required about 80 min. A second test took place 15 days later, using frequencies of 11 and 16 MHz, and covered the eastern part of the Straits. Radial-current measurements were interpolated to a common grid and combined within the overlapping region to produce the current vectors shown in Fig. 1 for the two test days. Some license has been taken in plotting the two measurements, made 15 days apart, on the same map, but only minor changes in these currents are expected in that time.

The map reveals the core of the Florida Current, which becomes the Gulf Stream off the US east coast. A maximum radar-derived current speed of 2.0 m s⁻¹ is recorded just off the east coast of Florida, a value consistent with *in situ* measurements in this region⁷. The northward (anticlockwise) flow at the western extremity of the map appears to be the eastern part of a semi-permanent feature sometimes called the Tortugas Gyre, centred at 84.5°W, 25°N in a concurrent TOPEX/Poseidon sea-surface-topography image. The southward flow that 'peels off' from the southeastern edge of the Florida Current at 80.5°W appears to be diverted to the south of the Cay Sal Bank, centred at 80°W, 23.8°N. The very small current values (with unresolved directions) on the shelf to the north of the Florida Keys are consistent with the shallow waters (and associated bottom friction) there. More frequent measurements would be required to remove any embedded tidal currents.

No concurrent *in situ* measurements were available for direct comparison with these radar measurements, underscoring the problem that the radar technique addresses. However, day-to-day consistency of single-radar current maps in the same region from the preceding and following days lends credibility to these features, none of which has been consistently and quantitatively observed using other techniques. Subsequent tests have validated the ROTHR current measurements by comparing them with a ship-borne acoustic Doppler current profiler traversing the Gulf of Mexico Loop Current (T.M.G., J.A.H. and R.A.L., unpublished data).

These two examples of high-resolution maps of ocean-surface currents illustrate the potential of this OTH radar pair for

mapping the surface circulation of the Intra-Americas Sea (IAS). Little is known about the circulation and its variability in and around the straits and channels of the West Indies, about the variable contributions to Florida Current transport by its subsidiary channels, and about the dynamics of the Gulf of Mexico Loop Current and its eddy-shedding process. Observations with enhanced space and time resolution of these and other poorly understood features could improve IAS circulation models (which are inadequately supported by observations) and lead to better understanding of how internal IAS dynamics modulate the western Atlantic boundary current and its poleward heat transport.

Because those specifications of OTH radars that apply to environmental monitoring are unclassified, other existing OTH radars that cover different ocean areas could be used for similar circulation studies. Our tests have shown that military radar

resources can be time-shared so as not to interfere with the radars' primary missions. As the strength of sea echoes observed with military OTH radars (which typically radiate 100 kW) is often 60 dB or more above atmospheric noise, consideration could even be given to designing low-power, low-cost OTH radars specifically for ocean monitoring. □

Received 24 July; accepted 8 December 1995.

1. *Proc. IEEE Fifth Working Conference on Current Measurement* (Institute of Electrical and Electronics Engineers, Piscataway, NJ, 1995).
2. Stewart, R. H. *Satellite Oceanography* 278–285 (Univ. California Press, Berkeley, CA, 1985).
3. Barrick, D. E., Evans, M. W. & Weber, B. *Science* **198**, 138–144 (1977).
4. Hammond, T. M. et al. *Continental Shelf Res.* **7**, 411–431 (1987).
5. Georges, T. M. & Harlan, J. A. *Eos* **76**, 146 (1995).
6. Georges, T. M. & Harlan, J. A. *IEEE Antennas and Propagation Mag.* **36**(4), 14–24 (1994).
7. Lee, T. N. et al. *J. Geophys. Res.* **100**(C5), 8607–8620 (1995).

Effect of depth-dependent viscosity on the planform of mantle convection

Hans-Peter Bunge*†, Mark A. Richards† & John R. Baumgardner*

* Institute of Geophysics and Planetary Physics, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA

† Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

LITHOSPHERIC plate motions at the Earth's surface result from thermal convection in the mantle¹. Understanding mantle convection is made difficult by variations in the material properties of rocks as pressure and temperature increase from the surface to the core. The plates themselves result from high rock strength and brittle failure at low temperature near the surface. In the deeper mantle, elevated pressure may increase the effective viscosity by orders of magnitude^{2–5}. The influence of depth-dependent viscosity on convection has been explored in two-dimensional numerical experiments^{6–8}, but planforms must be studied in three dimensions. Although three-dimensional planforms can be elucidated by laboratory fluid dynamic experiments^{9,10}, such experiments cannot simulate depth-dependent rheology. Here we use a three-dimensional spherical convection model^{11,12} to show that a modest increase in mantle viscosity with depth has a marked effect on the planform of convection, resulting in long, linear downwellings from the upper surface boundary layer and a surprisingly 'red' thermal heterogeneity spectrum, as observed for the Earth's mantle¹³. These effects of depth-dependent viscosity may be comparable to the effects of the plates themselves.

The notion of a weak zone in the upper mantle (the 'asthenosphere') is a recurrent and controversial theme in geodynamics. A variety of modern geophysical evidence suggests an upper mantle that is significantly weaker than the lower mantle. Such evidence includes post-glacial rebound¹⁴, the geoid^{15,16}, slow relative motions among volcanic hotspots⁷, geochemical mixing studies^{7,18} and long-term rotational dynamics¹⁹. Although depth resolution in such studies is coarse, most researchers have concluded that the average lower-mantle viscosity is between one and two orders of magnitude greater than that of the upper mantle, although some have argued for a smaller increase^{20–22}.

Recently published three-dimensional spherical convection models^{23,24} simultaneously include a number of depth-variable material properties (thermal expansivity, thermal conductivity, viscosity), mixed heating modes (internal and external) and compressibility effects. Here, instead, we examine a simple

model in order to focus on the effect of depth-dependent viscosity. Figure 1 shows models obtained assuming that the mantle is incompressible and that all material properties are constant except for viscosity, which may vary with depth (see Table 1).

We allow for uniform volumetric heating, to keep the numerical experiments simple and because it is thought that Earth's mantle is heated predominantly from within^{25–27}. Thermal density variations give rise to buoyancy forces that drive convection, but density variations are ignored otherwise (the 'Boussinesq' approximation). The vigour of convection is governed by the thermal Rayleigh number (a dimensionless buoyancy parameter) based on internal heating. The thermal boundary conditions are insulation at the core–mantle boundary (CMB) and constant temperature at the surface. Mechanical boundary conditions are free-slip.

Figure 1a shows the near-surface and interior temperature field for a calculation at a moderate Rayleigh number of $Ra = 1.6 \times 10^5$ (the Rayleigh number of the Earth's mantle is probably of the order of 10^8 or greater) and constant mantle viscosity. Convection is time-dependent, and the time snapshot shown (like others to follow) is representative of results after the calculation has reached quasi-equilibrium—over 15,000 time steps in this case. The blue regions are cold and the red regions are warm, corresponding to downwellings and upwellings, respectively. The planform in Fig. 1a is dominated by the sinking of the cold upper boundary layer at downwelling plumes. Upwellings are diffuse and there is no thermal boundary layer at the CMB, as expected for internally heated convection. The adjacent Fig. 1e

TABLE 1 Model parameters

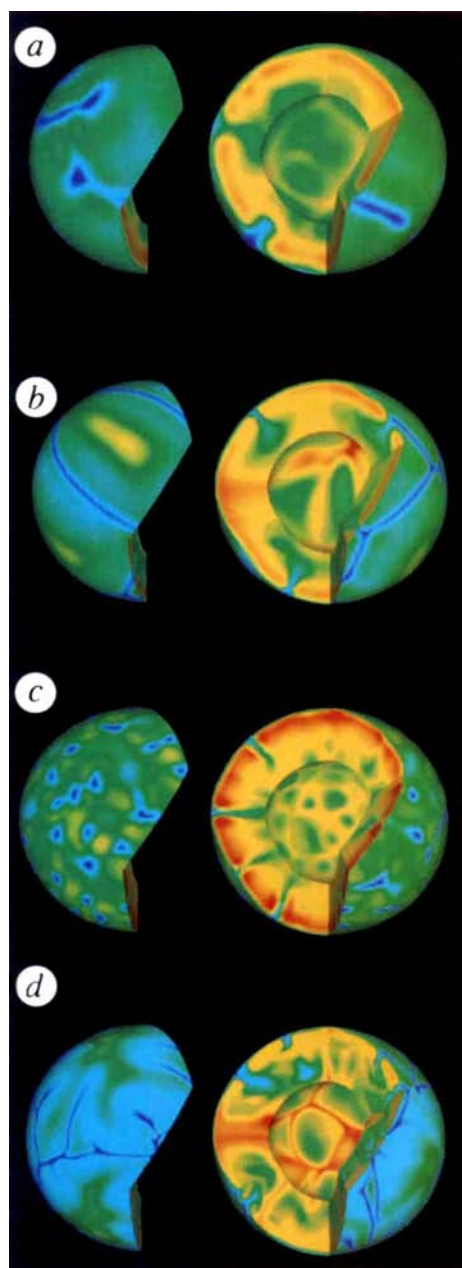
Outer shell radius	6,370 km
Inner shell radius	3,480 km
Reference density ρ_0	4,500 kg m ⁻³
Coefficient of thermal expansion α	2.5×10^{-5}
Gravitational acceleration g	10.0 m s ⁻²
Internal heating rate Q_{int}	1.0×10^{-12} W kg ⁻¹
Heat capacity	1.0×10^3 J kg ⁻¹ K ⁻¹
Thermal conductivity	4.0 W m ⁻¹ K ⁻¹
T , surface	1,060 K
Upper-mantle dynamic viscosity $\nu(a)$	1.7×10^{24} Pa s
Upper-mantle dynamic viscosity $\nu(b)$	5.8×10^{22} Pa s
Upper-mantle dynamic viscosity $\nu(c)$	5.8×10^{22} Pa s
Upper-mantle dynamic viscosity $\nu(d)$	7.0×10^{21} Pa s

Layered viscosity models

Lower-mantle dynamic viscosity $\nu(b, d)$ $30 \times \nu$ -upper mantle

Our numerical modelling code TERRA has been described elsewhere^{11,12}. The code solves for momentum and energy balance at infinite Prandtl number in a spherical shell. We use a finite element mesh of more than 10 million node points, providing an element resolution of roughly 50 km throughout the mantle.

FIG. 1 *a*, Temperature distribution for a convection calculation made for an incompressible, internally heated, isoviscous mantle with $Ra = 1.6 \times 10^5$. Blue is cold, and red is hot. The uppermost 250 km of the mantle is removed to show the temperature distribution beneath the boundary layer. Isolated downwelling plumes from the upper boundary layer dominate the planform. *b*, Same as *a* except that the viscosity of the upper mantle has been decreased by a factor of 30. The planform is dominated by long, linear downwelling sheets. *c*, Same as *a* except that the viscosity of the entire mantle has been decreased by a factor of 30. The planform is dominated by closely spaced downwelling plumes. *d*, Layered viscosity case as in *b* except that the viscosities of both the upper and lower mantle have been decreased by about a factor of 10. Linear sheets dominate the planform. *e-h*, Temperature power spectra from the incompressible convection calculations. The fully normalized spherical harmonic coefficients of the temperature variations in each layer in the model are squared and summed at each harmonic degree, and then averaged for each degree over the depth of the mantle. The power is normalized to the largest coefficient for each case.



shows the power spectrum of temperature heterogeneity averaged over the depth of the mantle. This spectrum is fairly 'red', with power concentrated at spherical harmonic degrees less than 15 and peaked at degree 3.

Figure 1*c* shows a similar calculation, except that the viscosity has been lowered by a factor of 30, resulting in more vigorous convection at a Rayleigh number of 4.8×10^6 . The planform is again dominated by downwelling plumes, but in this case the plumes are more closely spaced than in Fig. 1*a*. The spectrum (Fig. 1*g*) reflects this difference, showing a relatively 'blue' signature with power concentrated at harmonic degrees > 10 .

Figure 1*b* shows a model in which the lower mantle below 670 km depth has the higher viscosity of Fig. 1*a* and the upper mantle has the lower viscosity of the model of Fig. 1*c*. In other words, this is a hybrid calculation with a factor of 30 increase in viscosity with depth. This factor of 30 is chosen as a representative value from our previous work on modelling the geoid and other studies^{15,16}. In this hybrid calculation, the 'surface' Rayleigh number based on upper-mantle viscosity is 4.8×10^6 , whereas the Rayleigh number based on the arithmetic average of mantle viscosity is slightly over 5.7×10^5 .

The effects of this simple change in viscosity structure are considerable. First, the planform of Fig. 1*b* is completely transformed, with long, linear downwellings dominating convection in the upper mantle. Second, the spectrum (Fig. 1*f*) is even more 'red' than in the upper panel (Fig. 1*e*), with power concentrated from degrees 2 to 5. Both of these results are surprising. The further reddening of the spectrum from that of Fig. 1*e* occurs despite an increase in the average Rayleigh number. In isoviscous convection, heterogeneity spectra become less red as the Rayleigh number increases²⁸. More important, neither of the constant-viscosity calculations (Fig. 1*a* and *c*) shows any hint of the hybrid planform of Fig. 1*b*.

It is logical to ask whether these effects depend on the vigour of convection, or Rayleigh number. In the calculation of Fig. 1*d*, we have maintained the viscosity contrast of Fig. 1*b*, while increasing the Rayleigh number by almost an order of magnitude. Here the 'surface' Rayleigh number is 4.0×10^7 , and the 'average' Rayleigh number is 2.8×10^6 . The results of this calculation are even more dramatic than before. The viscosity increase transforms closely spaced downwelling plumes (Fig. 1*c*) into long, linear downwellings (Fig. 1*d*), and the heterogeneity spectrum is transformed

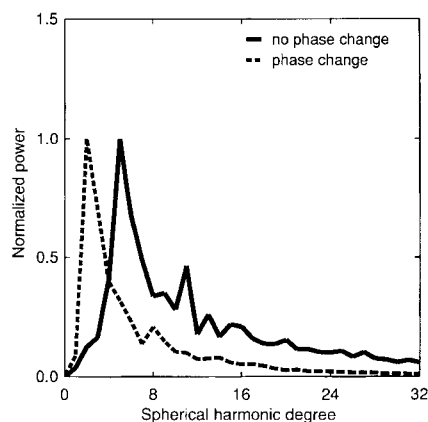


FIG. 2 Temperature power spectra for the model of ref. 23 without (solid) and with (dashed) an endothermic phase change at 670 km depth.

from very blue (Fig. 1g) to very red (Fig. 1h). Thus increasing the Rayleigh number seems to enhance the effect of viscosity layering, suggesting that the hybrid planforms and red heterogeneity spectra are basic properties of convection with layered viscosity.

Zhang and Yuen²⁹ have previously studied convection with depth-dependent viscosity. But they used predominantly basally heated models and focussed on hot upwellings from the CMB. They did not recognize the transition in the planform of purely internally heated convection, caused by the viscosity jump, from isolated cylindrical downwellings in the upper boundary layer to interconnected planar sheets.

In the hybrid models of Fig. 1b and d the typical surface velocities are much higher than for isoviscous convection with identical upper mantle viscosity. Thus the mantle achieves a larger aspect ratio by augmented lateral transport within the low-viscosity layer of the upper mantle. These qualitative remarks aside, we do not at present have a theoretical explanation for the hybrid planform observed. Two-dimensional stability analysis³⁰ does suggest a preference for greater planform lengthscales in layered viscosity convection, but we are not aware of any three-dimensional analyses.

The spectral reddening due to an increase in mantle viscosity is of great interest, because there is substantial evidence from seismic tomography and the geoid that the mantle heterogeneity spectrum is very red. The geoid has a strong peak at spherical

harmonic degrees 2–3 (ref. 31). Likewise global tomographic studies have consistently shown a peak in the spectrum of mantle seismic-velocity heterogeneity at degree 2 (refs 13,32). Although the spectra for layered viscosity models peak at spherical harmonic degrees 3–5 (Fig. 1f and h) rather than degree 2, their tendency to have greater low-degree structure than isoviscous models is clear. Combining such models with surface plates (and perhaps continents) should push the spectral peak toward lower harmonics³³.

In a recent study²³ it was concluded that modelling an endothermic phase transition at 670 km depth with a Clapeyron slope of $\gamma = -4 \text{ MPa K}^{-1}$ results in a shift of lateral heterogeneity from degrees 8–20 to degrees 2–6. But this phase-change-modulated calculation included substantial bottom heat flux (40 per cent from the core), which was noticeably smaller for their case with $\gamma = 0$. We have reproduced this strongly basally heated model (including a 10-fold increase in lower-mantle viscosity) and we also obtain a red spectrum as shown in Fig. 2 (dashed curve). But we observe a strong peak at degree 5 if we set $\gamma = 0$ (solid curve in Fig. 2) in this model. Apparently the 10-fold increase in radial viscosity, when combined with bottom heat flux, is very effective in promoting long-wavelength heterogeneity.

Our calculations indirectly address two other problems of geodynamics. First, the slow relative motions of hotspots may be accounted for by an increase in lower-mantle viscosity by a factor of 30–100 (ref. 17). Indeed, our models show that deep-mantle horizontal advection velocities decrease roughly in proportion to the logarithm of the ratio of upper/lower-mantle viscosity, as shown previously for two-dimensional models⁷. Second, a viscosity contrast of about two orders of magnitude may also be sufficient to isolate the lower mantle so as to preserve the ancient geochemical heterogeneities for which there is evidence from basalts erupted at ocean islands or hotspots^{7,18}.

Other than lacking plates, the main shortcoming of our models is the absence of horizontal viscosity variations due to thermal (or stress) variations. It is not clear, aside from the piecewise rigidity of the upper boundary layer itself, how strongly lateral viscosity variations in the deep mantle affect convection^{8,34}, or whether our main conclusions would be changed by modelling them. This question needs to be pursued in future studies. In addition, we have neglected heat flux from the core, which is certainly an approximation. We have, however, examined models with 20 per cent basal heating, and the results are not substantially different. We have also assumed that the mantle is of uniform composition, whereas a significant increase in intrinsic mantle density with depth could, of course, affect the style of convection. □

Received 11 August; accepted 20 December 1995.

1. Turcotte, D. L. & Oxburgh, E. R. *J. Fluid Mech.* **28**, 29–42 (1967).
2. Sammis, C. G., Smith, J. C. & Schubert, G. *J. geophys. Res.* **86**, 10707–10718 (1981).
3. Sammis, C. G., Smith, J. C., Schubert, G. & Yuen, D. A. *J. geophys. Res.* **82**, 3747–3761 (1977).
4. Ringwood, A. E. *J. Geol.* **90**, 611–643 (1982).
5. Jeanloz, R. & Thompson, A. B. *Rev. Geophys. Space Phys.* **21**, 51–74 (1983).
6. Foster, T. D. *J. geophys. Res.* **74**, 685–693 (1969).
7. Gurnis, M. & Davies, G. F. *Geophys. J. R. astr. Soc.* **85**, 523–541 (1986).
8. Christensen, U. R. *Geophys. J. R. astr. Soc.* **77**, 343–384 (1984).
9. Busse, F. H. & Whitehead, J. A. J. *Fluid Mech.* **47**, 305–320 (1971).
10. White, D. B. *J. Fluid Mech.* **91**, 247–286 (1988).
11. Baumgardner, J. R. *J. stat. Phys.* **39**, 501–511 (1985).
12. Bunge, H.-P. & Baumgardner, J. R. *Comput. Phys.* **9**, 207–215 (1995).
13. Su, W. J. & Dziewonski, A. M. *Nature* **352**, 121–126 (1991).
14. Nakada, M. & Lambeck, K. *Geophys. J.* **96**, 497–517 (1989).
15. Hager, B. H. & Richards, M. A. *Phil. Trans. R. Soc. Lond.* **328**, 309–327 (1989).
16. Ricard, Y., Richards, M. A., Lithgow-Bertelloni, C. & Le Stunff, Y. *J. geophys. Res.* **98**, 21895–21909 (1993).
17. Richards, M. A. in *Glacial Isostasy, Sea Level and Mantle Rheology* (eds Sabadini, R., Lambeck, K. & Boschi, E.) 571–588 (Kluwer, Dordrecht, 1991).
18. Gurnis, M. & Davies, G. F. *Geophys. Res. Lett.* **85**, 541–544 (1986).
19. Richard, Y., Sabadini, R. & Spada, G. *J. geophys. Res.* **97**, 14223–14236 (1992).
20. Cathles, L. M. *The Viscosity of the Earth's Mantle* (Princeton Univ. Press, 1975).

21. Peltier, W. R. *A. Rev. Earth planet. Sci.* **9**, 119–225 (1981).
22. Forte, A. M. & Peltier, W. R. *J. geophys. Res.* **96**, 20131–20159 (1991).
23. Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A. & Schubert, G. *Nature* **361**, 699–704 (1993).
24. Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A. & Schubert, G. *J. geophys. Res.* **99**, 15877–15901 (1994).
25. Wasserburg, G. J., MacDonald, G. J. F., Hoyle, F. & Fowler, W. A. *Science* **143**, 465–467 (1964).
26. Davies, G. F. *J. geophys. Res.* **93**, 10467–10480 (1988).
27. Sleep, N. H. *J. geophys. Res.* **95**, 6715–6736 (1990).
28. Jarvis, G. T. & Peltier, W. R. *J. geophys. Res.* **91**, 435–451 (1986).
29. Zhang, S. & Yuen, D. A. *Earth planet. Sci. Lett.* **132**, 157–166 (1995).
30. Buffett, B. A., Gable, C. W. & O'Connell, R. J. *J. geophys. Res.* **99**, 19885–19900 (1994).
31. Lerch, F. J., Klosko, S. M. & Patch, G. B. *NASA Tech. Memo* 84,986 (1983).
32. Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y. *Phys. Earth planet. Inter.* **59**, 294–328 (1990).
33. Gurnis, M. & Zhong, S. *Geophys. Res. Lett.* **18**, 581–584 (1991).
34. Tackley, P. J. *Geophys. Res. Lett.* **20**, 2187–2190 (1993).

ACKNOWLEDGEMENTS. We thank P. Tackley for comments on the manuscript and for assistance in reproducing the model results in refs 23 and 24. We also thank U. Christensen, R. O'Connell and G. Glatzmaier for reviews. This work was supported by the National Science Foundation and the Institute of Geophysics and Planetary Physics at Los Alamos National Laboratory. Computing time was provided by the Advanced Computing Laboratory (ACL) at Los Alamos. H.-P.B. thanks the members of the ACL.

Oldest fossil record of gliding in rodents

G. Storch*, B. Engesser† & M. Wuttke‡

* Forschungsinstitut Senckenberg, Senckenberg-Anlage 25, D-60325 Frankfurt am Main, Germany

† Naturhistorisches Museum, Augustinergasse 2, CH-4001 Basel, Switzerland

‡ Landesamt für Denkmalpflege Rheinland-Pfalz, Referat Erdgeschichtliche Denkmalpflege, Am Obstmarkt 13, D-55126 Mainz, Germany

EOMYIDAE is an extinct family of rodents with a wide distribution in North America, Europe and Asia¹⁻³. Of the modern rodent groups, eomyids are most closely related to New World pocket mice (heteromyids) and pocket gophers (geomyids)⁴. Eomyids occurred from the late Eocene through the Pliocene, spanning a time period of about 40 million years. From Europe alone, 11 genera and about 50 species have been recognized. The time of greatest diversity was the late Oligocene and early Miocene when eomyids dominated many small mammal faunas^{1,5}. Their fossil record, however, consists almost exclusively of teeth, with the postcranial skeleton being virtually unknown. Here we present a complete and extraordinarily well-preserved eomyid. Its soft body outline strongly suggests the existence of gliding membranes (Figs 1, 2). Thus eomyids are the fourth family of rodents (in addition to squirrels (Sciuridae), scaly-tailed flying squirrels (Anomaluridae), and dormice (Gliridae)) with representatives capable of gliding locomotion.

The new eomyid fossil is from Enspel near Bad Marienberg in the Westerwald, state of Rheinland-Pfalz, Germany. The Enspel

Fossilagerstätte is a former lake of late Oligocene age. The lake deposits, bituminous volcanoclastic shales which are episodically intercalated with tuffs, are more than 100 m thick and overlain by a thick basalt. Excavations since 1990 by the Department for the Protection and Conservation of Cultural Monuments of Rheinland-Pfalz have yielded numerous plant, invertebrate and vertebrate fossils, including insects, fish, anurans, caudatans, crocodiles and mammals.

The Enspel eomyid is identified as *Eomys quercyi* Comte & Vianey-Liaud, 1987. The teeth were extracted before the fossil was finally transferred to an artificial matrix of epoxy-resin (Fig. 3). They exhibit diagnostic specific features like strikingly short mesolophids on P₄-M₃, strongly reduced M³, and poorly developed first syncline on M² (ref. 6). *E. quercyi* is currently only known from the late Oligocene European Paleogene Mammal Zone MP 28. Based on correlations with the paleomagnetic time scale, zone MP 28 has an age of around 25.8 Myr (ref. 7).

The articulated eomyid skeleton is in a relaxed pose, as is often seen in drowned mammals. However, in contrast with other fossils of drowned rodents, which are usually preserved on their sides and have the right and left extremities in a relaxed parallel posture, the specimen from Enspel is lying on its back with its thighs spread out. Therefore we can expect that it differed by some particular feature of the body from a generalized rodent. Head and thorax are rotated somewhat laterally, and the distal segments of the left limbs are folded towards the right side of the body. The distal part of the tail is missing and the symphysis of the mandible, both wrists, and one knee joint are dislocated to some extent. The carcass shows an intact, continuous, soft body outline with preserved imprints of hairs and no signs of major distension or decay. Thus the animal must have drowned and quickly sunk to the bottom of the lake before gases of putrefaction collected in the abdominal cavity, and it

FIG. 1 Skeleton of *Eomys quercyi* with preserved soft body outline from the late Oligocene of Enspel, Westerwald, Germany (Naturhistorisches Museum, Mainz, specimen 5803/G3S16, found in 1992). Length of head and body, about 10 cm; skull length, 3 cm. The carcass shows thick fur along its left side (above) and the sparsely haired, thin gliding membrane along the right side (below). Even delicate structures such as the rib cage and the sternal apparatus are preserved in their natural articulation, and we can thus exclude the possibility that the outline of the soft body is due to post-mortem crashing. The equally well-preserved rodents from Messel near Frankfurt, Germany, did not possess gliding membranes and are fossilized in a normal pose, lying on their side with their feet folded¹⁵.



FIG. 2 Close up of the forelimb of the specimen shown in Fig. 1. The sparsely haired gliding membrane extends from the neck to the dislocated wrist and from there posteriorly; its outer margin is straight and well delimited. A rod extending from the elbow proximally does not accord in morphology and orientation with adjacent elements, such as ribs or sternal segments. It is too long and strong to represent the tendon of the triceps muscle, and seems to be a calcar for support of the patagium. It points in the opposite direction to that expected for a calcar, yet it is located within a narrow zone of disruption (light in the photograph) and may be thus displaced medially together with adjoining pieces of the membrane.



FIG. 3 Upper (a) and lower (b) cheek teeth of *Eomys quercyi* from Enspel. With the exception of the two premolars, the teeth are heavily worn and indicative of an old individual. In addition to *Eomys quercyi*, two more eomyid species are known by their dentitions from the Enspel site. They come, however, from fossil owl pellets and are represented only by isolated teeth and scraps of postcranial bones. Therefore, nothing can be concluded thus far about the particular biological strategies of these closely related taxa.



FIG. 4 Reconstruction of *Eomys quercyi* from Enspel during gliding locomotion. The distal part of the tail is not preserved and its reconstructed length is about 10 cm. Small gliders such as *Eomys quercyi* have low wing loadings, and are thus adapted to a manoeuvrable and slow 'flight' in densely forested areas where there is little air turbulence. Abundant fossils of leaves, fruits and seeds prove the existence of a thick mesophytic forest around where the rodent was found, and this forest type was widespread in western and central Europe during the late Oligocene¹⁶.



was buried while the decomposition of the hind limbs was still at an early stage⁸.

Of special interest is the clearly defined contour of the soft body. Because of the slight rotation, the outlines of the left and right sides of the carcass differ markedly. On the left side, the dense and rather long-haired pelage extends from the muzzle to the rump; long vibrissae and parts of the external ear are also visible. The outline of the right side, however, is formed by a thin, sparsely haired, and laterally expanded membrane which is spread out between the neck, wrist, ankle and proximal tail vertebrae. The boundary between the membranous skin and the compact, densely haired body is particularly obvious in front of the foreleg. The outline along the trunk is straight and parallel to the long axis of the body. We conclude that the membrane acted as a patagium and *Eomys quercyi* was a glider (Fig. 4). Its preservation in a dorsoventral position results obviously from the presence of the gliding membranes. This discovery contradicts the supposition, based on the distribution and abundance of extant gliders, that the evolution of gliding and the occupation of the gliding niche were linked to rich tropical-forest ecosystems and that particular tropical vegetation structures figured prominently in this connection^{9,10}.

Glissant locomotion has developed by parallel evolution in four rodent families. Among extant rodents, these are flying squirrels and scaly-tailed flying squirrels. The patagia are supported in many forms by cartilaginous struts or calcars, which originate from the carpus in sciurids and from the olecranon process in anomalurids. Presumably, *Eomys quercyi*, like anomalurids, had a long strut from the elbow; a rod extends from the right olecranon of the Enspel specimen and seems to represent the medially displaced calcar. Flying squirrels are characterized by the elongation of the distal elements of the forelimbs (ratio of radius : humerus = 0.99–1.13 (ref. 11), which is not found in either *Anomalurus* (0.90) or *E. quercyi* (0.85). The dormouse family also includes a glider: its living members are non-gliding, mainly arboreal animals, the sole glissant species being *Glirulus* aff. *lissiensis* Huguency & Mein, 1965 from the late Miocene of the Ardèche, France¹². The fossil dormouse is similar to *E. quercyi* from Enspel in the dorsoventral position of the carcass and the imprint of the patagium. Note that it is referred by dental characters to the same genus as the living non-gliding Japanese species, *Glirulus japonicus* (Schinz, 1845).

Accordingly, we cannot infer gliding adaptations for all eomyids, or even all species of the genus *Eomys*, from this specimen of *E. quercyi*. *Eomys* is much too frequent and diverse in many European localities for one to assume that all species were gliders. Previous postcranial evidence of eomyids consisted of parts of a humerus, pelvis, femur, tibia and calcaneum from the North American middle Oligocene¹³, and suggested a cursorial locomotion as well as 'an aberrant adaptation'¹⁴. The highly specialized locomotor strategy of *E. quercyi*, however, could not have been inferred from the living members of the clade (heteromyids and geomyids) or from previously available fossils of the family Eomyidae. □

Received 25 July; accepted 13 November 1995.

- Engesser, B. *Schweiz. Paläont. Abh.* **112**, 1–144 (1990).
- Korth, W. W. *The Tertiary Record of Rodents in North America* (Plenum, New York, 1994).
- Qiu, Zh.-d. *Natn. Sci. Mus. Monogr., Tokyo* **8**, 49–55 (1994).
- Fahlbusch, V. in *Evolutionary Relationships among Rodents. A Multidisciplinary Analysis* (ed. Luckett, P. W. & Hartenberger, J.-L.) 617–629 (Plenum, New York, 1985).
- Alvarez Sierra, M. A. *Scripta geol.* **86**, 1–207 (1987).
- Comte, B. & Vianey-Liaud, M. *C. R. Acad. Sci., Paris* **304**, 951–954 (1987).
- Burbank, D. W. et al. *Eclogae geol. Helv.* **85**, 399–431 (1992).
- Wuttke, W. in *Messel. An Insight into the History of Life and of the Earth* (ed. Schaal, S. & Ziegler, W.) 263–275 (Clarendon, Oxford, 1992).
- Emmons, L. H. & Gentry, A. H. *Am. Nat.* **121**, 513–524 (1983).
- Dudley, R. & DeVries, P. *Biotropica* **22**, 432–434 (1990).
- Thomington, R. W. & Heaney, L. R. *J. Mamm.* **62**, 101–114 (1981).
- Mein, P. & Romaggi, J.-P. *Geobios* **13**, 45–50 (1991).
- Cope, E. D. *Rep. U.S. geol. Surv. Terr.* **3**, 1–1009 (1884).
- Wood, A. E. *Trans. Am. phil. Soc.* **28**, 155–269 (1937).
- Koenigswald, W. von, Storch, G. & Richter, G. in *Messel. An Insight into the History of Life and of the Earth* (ed. Schaal, S. & Ziegler, W.) 217–222 (Clarendon, Oxford, 1992).
- Mai, D. H. & Walther, H. *Abh. staatl. Mus. Mineral. Geol., Dresden* **38**, 1–230 (1991).

ACKNOWLEDGEMENTS. We thank Jin Meng and J. J. Hooker for comments. P. Schäfers skillfully prepared the specimen described herein.

A canine distemper virus epidemic in Serengeti lions (*Panthera leo*)

Melody E. Roelke-Parker^{*†}, Linda Munson^{‡##},
Craig Packer^{*§}, Richard Kock^{||},
Sarah Cleaveland^{*¶#}, Margaret Carpenter[☆],
Stephen J. O'Brien[☆], Andreas Pospischil^{**},
Regina Hofmann-Lehmann^{††}, Hans Lutz^{††},
George L. M. Mwamengele^{‡‡}, M. N. Mgasa^{‡‡},
G. A. Machange^{§§}, Brian A. Summers^{|||}
& Max J. G. Appel^{¶¶}

^{*} Serengeti Wildlife Research Institute, Tanzania National Parks, Arusha, Tanzania

[†] Messerli Foundation, Zurich, Switzerland

[‡] Department of Pathology, College of Veterinary Medicine, University of Tennessee, Knoxville, Tennessee 37901, USA

[§] Department of Ecology, Evolution, and Behavior, University of Minnesota, St Paul, Minnesota 55108, USA

^{||} Kenya Wildlife Service, Nairobi, Kenya

[¶] Institute of Zoology, London NW1 4RY, UK

[#] London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

[☆] Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21702, USA

^{**} Institute of Veterinary Pathology, and ^{††} Department of Veterinary Internal Medicine, University of Zurich, 8057 Zurich, Switzerland

^{‡‡} Department of Veterinary Pathology, Sokoine University of Agriculture, Morogoro, Tanzania

^{§§} Veterinary Investigation Center, Arusha, Tanzania

^{|||} Department of Pathology, and ^{¶¶} James Baker Institute of Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

CANINE distemper virus (CDV) is thought to have caused several fatal epidemics in canids within the Serengeti–Mara ecosystem of East Africa, affecting silver-backed jackals (*Canis mesomelas*) and bat-eared foxes (*Otocyon megalotis*) in 1978 (ref. 1), and African wild dogs (*Lycan pictus*) in 1991 (refs 2, 3). The large, closely monitored Serengeti lion population^{4,5} was not affected in these epidemics. However, an epidemic caused by a morbillivirus closely related to CDV emerged abruptly in the lion population of the Serengeti National Park, Tanzania, in early 1994, resulting in fatal neurological disease characterized by *grand mal* seizures and myoclonus; the lions that died had encephalitis and pneumonia. Here we report the identification of CDV from these lions, and the close phylogenetic relationship between CDV isolates from lions and domestic dogs. By August 1994, 85% of the Serengeti lion population had anti-CDV antibodies, and the epidemic spread north to lions in the Maasai Mara National reserve, Kenya, and uncounted hyaenas, bat-eared foxes, and leopards were also affected.

In early 1994, six lions in the Serengeti National Park, Tanzania were observed with *grand mal* seizures, and three other lions developed facial and forelimb myoclonus (recurrent twitching). Additional lions were noted to be disoriented, ataxic and profoundly depressed. Between January and March 1994, 11 lion carcasses were found, representing a dramatic increase in mortality from previous years and indicating that a serious epidemic was emerging.

To investigate the epidemic, tissue and serum samples were obtained from 23 lions that died or were killed in a moribund state, 13 live lions with obvious signs of disease, and 72 apparently healthy, anaesthetized lions. Sera from 111 healthy lions sampled between 1984 and 1994 (refs 6–8) were also analysed to track previous exposure of this population to viruses. Tissues from 19

^{##} To whom correspondence should be addressed.

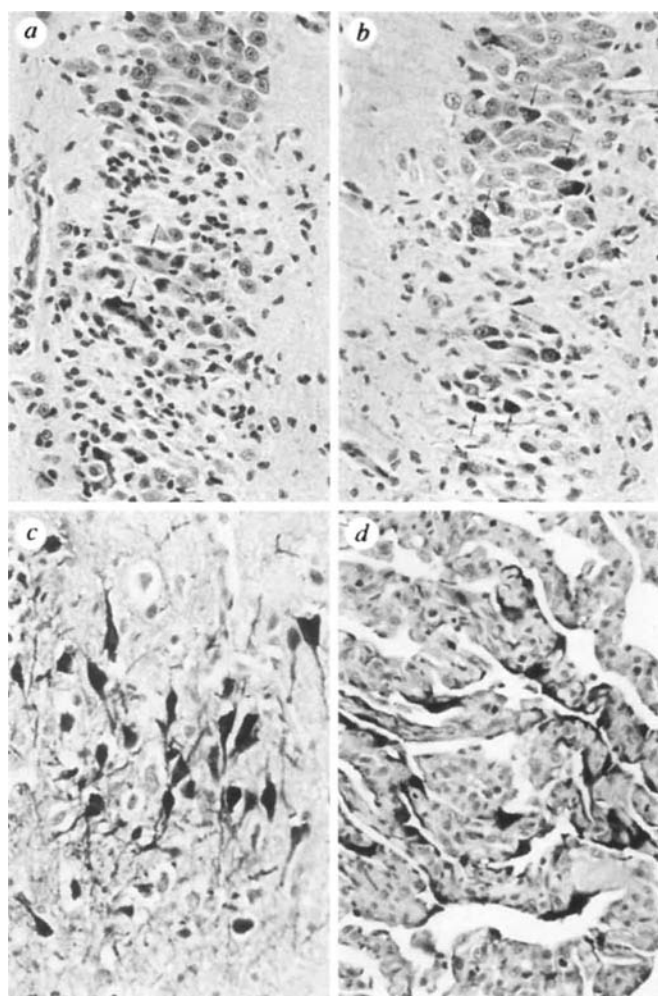


FIG. 1 Characteristic histopathological changes and intralesional immunoreactive viral nucleocapsid antigens in tissues of African lions infected with canine distemper virus. *a*, Haematoxylin- and eosin-stained section of the brain from a lion with CDV encephalitis ($\times 350$). The dentate gyrus in the hippocampus has neuronal loss, dense gliosis and occasional multinucleated giant cells (arrows). *b*, Immunohistochemical identification by monoclonal antibody N3.991 against CDV-nucleocapsid protein in a similar area to *a*. Several neurons contain CDV antigen in the cytoplasm (some indicated by arrows) ($\times 350$). *c*, Immunohistochemical identification of CDV nucleocapsid in a parahippocampal gyrus ($\times 350$). Many neurons are strongly positive for CDV antigen in perikarya and neurites. *d*, Immunohistochemical identification of CDV nucleocapsid proteins in lung with interstitial pneumonia, characterized by type II pneumocyte hyperplasia and alveolar septal thickening ($\times 350$).

METHODS. Tissues were fixed in 10% formalin, embedded in paraffin, sectioned at 5–7 μ m and stained with haematoxylin and eosin. Immunohistochemical procedures were applied to sections of formalin-fixed, paraffin-embedded tissues that had typical CDV lesions on stained sections to demonstrate intralesional viral antigens¹⁰. Paraffin-embedded tissues had the paraffin removed, were treated to remove endogenous peroxidase, then incubated with a mouse monoclonal antibody to a CDV-N protein (MAB N3.991)¹¹, or with rabbit primary polyclonal antibody raised with the Rockborn strain of CDV. A commercial avidin-biotin kit was used to identify sites of monoclonal antibody binding to tissues, and a commercial peroxidase-antiperoxidase kit was used for polyclonal antibodies, then tissue sections were counterstained with Gill's haematoxylin, dehydrated and mounted with Permount. Negative controls were duplicate sections stained using a monoclonal antibody for influenza virus replacing CDV monoclonal antibodies, and positive controls were brain sections from a confirmed case of CDV in a domestic dog.

dead lions, examined by histopathology, had either encephalitis, interstitial pneumonia, and/or lymphocytic depletion in lymph nodes and spleen (Fig. 1; Table 1). Rare multinucleated syncytia and intranuclear and/or intracytoplasmic viral inclusions characteristic of morbilliviral infection⁹ were also found in these lions (Table 1). Because these lesions were seen in zoo cats in the United States in the 1991 and 1992 canine distemper epidemics¹⁰, we used monoclonal¹¹ and polyclonal CDV antibodies to confirm that CDV nucleocapsid antigens were present in affected tissues. We then tested all available lion sera for neutralizing antibody titres to CDV^{10,12}, and found that 63 of the 72 apparently healthy lions and 8 of the 11 sick lions sampled in 1994 had CDV titres (Fig. 2a). We isolated CDV from the cerebrospinal fluid of one lion cub with *grand mal* seizures that subsequently died with CDV encephalitis¹³. The monoclonal antibody-binding pattern of this virus was compared to binding patterns of viruses isolated from: a bat-eared fox (*O. megalotis*), a spotted hyaena (*Crocuta crocuta*), and a domestic dog that died during the 1994 Serengeti epidemic; with CDV isolated from a lion during the 1992 California epidemic; with CDV isolated from a raccoon (*Procyon lotor*) and fox (*Vulpes vulpes*); and with a virulent CDV (A75-17, 1975) and two strains of attenuated CDV from domestic dogs (Rockborn 1958 and Onderstepoort 1948). Monoclonal antibodies were donated by C. Orvell (Huddinge, Sweden), and were directed against viral nucleoprotein (N), polymerase (P), fusion glycoprotein (F), and haemagglutinin glycoprotein (H). Lymphocytes infected with the Serengeti lion virus bound the same group of CDV monoclonal antibodies as cells infected with these other viruses, suggesting that the lion morbillivirus was CDV. To characterize further the lion morbillivirus, the genomic sequences of the CDV P gene, a conserved region of the virus, were amplified from buffy-coat

lymphocytes of two lions with neurological signs. Phylogenetic relationships between the P-gene sequences of the two RT-PCR-derived lion CDV and the P-gene sequences of other morbilliviruses were then examined. These analyses (Fig. 3) indicated that Serengeti lion CDV virus is closely related to the Onderstepoort strain of canine distemper virus isolated from a domestic dog in South Africa. The statistics of the epidemic are given in Table 1.

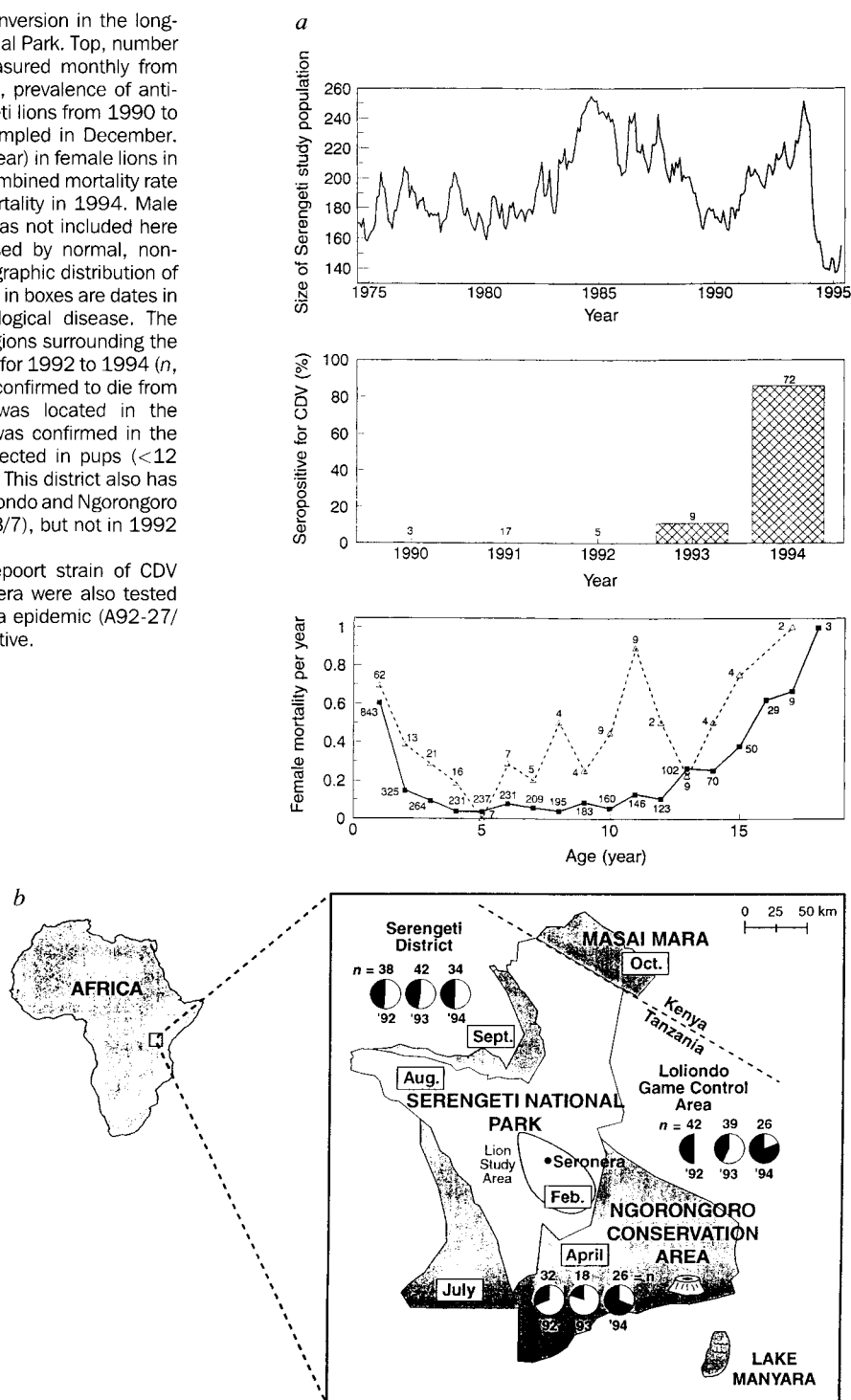
Later in 1994, CDV-infected lions were identified in the northern and western areas of the Serengeti National Park and in the Maasai Mara National Reserve in Kenya (Fig. 2b). Of 54 apparently healthy lions sampled from October 1994 to March 1995 in the Mara and neighbouring areas of Kenya, 23 had high serum titres of CDV antibodies, suggesting recent exposure.

During 1994, 39 CDV-associated lion deaths were documented, but this is probably an underestimation of true mortality statistics because most of the Serengeti lion population outside the Seronera study area is not under close observation. The overall lion population in the Serengeti ecosystem, estimated at 3,000 before the outbreak, is now estimated at 2,000. In 1994 and 1995, CDV-related deaths were also confirmed in seven spotted hyaenas (*C. crocuta*) by pathology and virology. Because CDV is historically not pathogenic in lions or hyaenas^{9,14}, the emerging biotype of CDV has apparently extended its host range. During the epidemic, CDV-induced disease also was confirmed in two bat-eared foxes, and CDV-like neurological disease was observed in a common jackal (*Canis aureus*) and two silver-backed jackals, indicating that the Serengeti CDV biotype conserves its pathogenicity for canids.

The magnitude of this epidemic may be explained in part by the lion population being immunologically naive to CDV when it was introduced in 1994. All but one of the 34 lions sampled between 1990 and 1993 were seronegative, and the seropositive lion was

Fig. 2 *a*, Temporal patterns of mortality and seroconversion in the long-term study population of lions in the Serengeti National Park. Top, number of lions resident in the long-term study area, measured monthly from August 1974 to February 1995 (refs 4, 5). Middle, prevalence of anti-canine distemper virus (CDV) antibodies¹⁰ in Serengeti lions from 1990 to 1994. The single seropositive lion in 1993 was sampled in December. Bottom, age-specific mortality (proportion dying per year) in female lions in the study population. The solid line represents the combined mortality rate from 1966 to 1993; the broken line represents mortality in 1994. Male mortality was comparable to female mortality, but was not included here because male disappearances could also be caused by normal, non-disease-related dispersals^{4,5}. *b*, Chronology and geographic distribution of CDV disease in the Serengeti ecosystem. The months in boxes are dates in 1994 when lions were first observed with neurological disease. The proportions of CDV-seropositive domestic dogs in regions surrounding the Serengeti National Park are represented by pie charts for 1992 to 1994 (*n*, number of dogs tested). The domestic dog that was confirmed to die from canine distemper disease in September 1994 was located in the Ngorongoro Conservation Area. Endemic infection was confirmed in the Serengeti district where CDV seropositivity was detected in pups (<12 months old) in each year of the study (1992–1994). This district also has one of the highest dog population densities. In the Loliondo and Ngorongoro areas, seropositivity was detected in pups in 1994 (3/7), but not in 1992 (0/6) or 1993 (0/17).

METHODS. Sera were tested against the Onderstepoort strain of CDV adapted to Vero cells as described previously¹². Sera were also tested against CDV isolated from a lion during the California epidemic (A92-27/20). Log titres of 1.0 or greater were considered positive.



sampled in December of 1993 (Fig. 2a). Although this abrupt seroconversion indicates a recent exposure of these lions to CDV, analysis of sera from 77 lions in the region from 1984 to 1989 disclosed that 22 lions had antibodies to CDV (none of these previously tested lions was alive during the 1994 epidemic), indicating that lions in the Serengeti ecosystem had previously encountered CDV without an increase in disease-related mortality. To determine if the high lion mortality during the 1994 CDV epidemic was due to co-infection with another viral pathogen, we compared the prevalences of serum antibody titres to feline immunodeficiency virus (FeIV)¹⁵, feline parvovirus (FePV)¹⁶, feline herpesvirus 1 (FeHV1), feline corona virus (FeCoV), and feline calici virus (FeCV) between lions with CDV disease (6/10 FeIV+; 3/5 FePV+; 6/6 FeHV+; 2/6 FeCoV+; 1/6 FeCV+) and

healthy CDV-seropositive lions (15/16 FeIV+; 11/12 FePV+; 13/13 FeHV+; 8/13 FeCoV+; 8/13 FeCV+). The discordance between CDV disease and antibodies to other viruses (Fisher's exact test: FeIV, $P > 0.055$, FePV, $P > 0.19$; FeCoV, $P > 0.35$; FeCV, $P > 0.14$) fails to support a role for these viruses as necessary cofactors in CDV morbidity.

CDV infections in equatorial African wildlife usually occur as periodic epidemics because environmental factors limit viral persistence outside susceptible carnivore hosts, and these hosts usually succumb or rid themselves of virus⁹. CDV persists in dense populations of domestic dogs because pups provide a constant reserve of susceptible hosts. The source of CDV in the Serengeti epidemic was probably the domestic dogs of the local villages adjacent to the National Park. CDV seroprevalance increased in

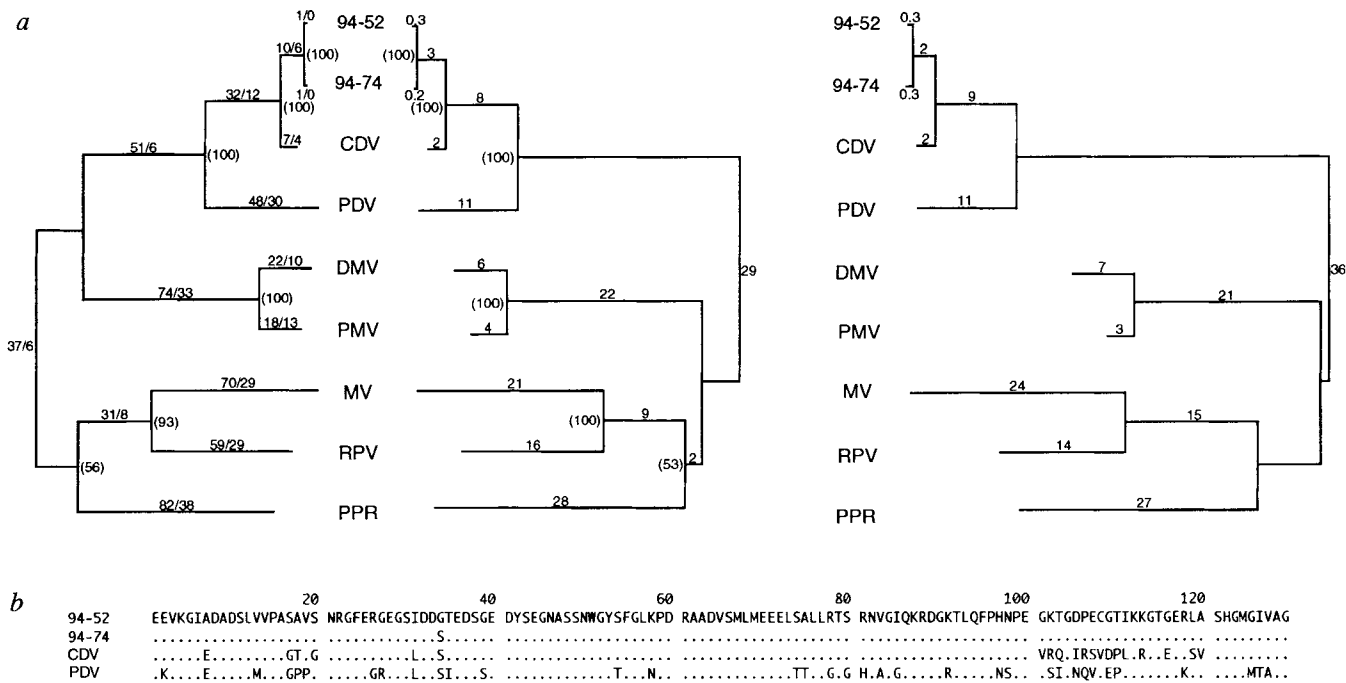


FIG. 3 Sequences and phylogenetic analysis of 389 bp from lion morbillivirus P genes. *a*, Viral RNA fragments of 429 bp were amplified by RT-PCR from RNA of thymus and white blood cells isolated from two lions with neurological signs. Oligonucleotide primers were synthesized based on conserved regions of the CDV phosphoprotein (P) gene as described previously¹⁸. PCR product fragments were cloned and sequenced, and the sequences aligned with P-gene sequences from other morbilliviruses. Derived P-gene sequences from two lions were 99% identical. A phylogenetic analysis using cladistic, phenetic and maximum-likelihood methods revealed a close relationship between the lion morbillivirus and the Onderstepoort strain of CDV, the sequences having 95% nucleotide identity. PAUP maximum parsimony tree¹⁹ (left), with branch lengths equal to the number of nucleotide substitutions. Numbers shown on branches are

branch length and number of homoplasies, in that order. Tree length, 543; consistency index, 0.790. PHYLIP neighbour-joining tree²⁰ (centre) generated using a transition/transversion ratio of 0.84, with branch lengths equal to the percent of nucleotide differences. The number of bootstrap iterations (out of 100) which support the nodes are shown in parentheses. PHYLIP maximum-likelihood tree (right) generated with a transition/transversion ratio of 0.84 (in likelihood, -2561.52597; 155 trees examined). Branch lengths are the expected number of substitutions per site $\times 100$. 94-52-10 and 94-74 are sequences from Serengeti lions. CDV, canine distemper (Genbank X51869); PDV, phocine distemper virus (Genbank X75960); MV, measles virus (Genbank X16569); RPV, rinderpest virus (Genbank X68311); DMV, dolphin morbillivirus; PMV, porpoise morbillivirus; and PPR, peste des petits ruminants¹⁸. *b*, Translated amino-acid sequence of lion morbillivirus compared to CDV and PDV.

TABLE 1 Summary of CDV disease in Serengeti lions	
Criteria indicating CDV infection	Number of affected lions
Neurological signs of CDV*	
Seizures	12
Myoclonus	15
Other neurological signs	27
Lion mortalities	
Carcasses recovered	23
Disappearances from observed population of 250 lions	31
CDV seroprevalence†	71 of 83
Lions with histopathological lesions of CDV	18 of 19
Lions with viral inclusions in tissues	14 of 19
Lions with CDV nucleocapsid proteins in tissues†	14 of 19
Lions with CDV isolated†	1 of 7
Lions with CDV RNA obtained by RT-PCR‡	2

* Number designates lions observed during 1994. Other neurological signs included ataxia, disorientation, profound depression or stupor, hyperaesthesia, and inappropriate behavioural responses.
† Immunohistochemistry and viral isolation were performed by methods published previously^{10,12}. Buffy-coat isolation for viral cultures and molecular analyses were derived from 20 ml blood taken from anaesthetized lions.
‡ Methods are described in Fig. 3 legend.

these dogs between 1991 and 1993, preceding the lion epidemic (Fig. 2*b*), and CDV encephalitis was confirmed by histopathology in a domestic dog in the Ngorongo Crater region in 1994. The precise route of CDV transmission to the lions is unclear, because direct dog-to-lion contact is unlikely for most of the Serengeti ecosystem. A more probable route is by spotted hyaenas, which range among human dwellings and travel long distances within the Park¹⁷. Nomadic lions could also contribute to viral dissemination. High densities of these susceptible carnivores at kill sites would then provide an ideal environment for CDV amplification and transmission.

Most of the lion deaths in the Serengeti National Park occurred between January and September 1994, and mortality rates have subsequently returned to previous levels (Fig. 2*a*). Although this CDV epidemic claimed approximately 30% of the Serengeti and Mara lions, the impact on other carnivore species is unknown. Less dense populations of endangered species, such as cheetahs or wild dogs, are a clear cause for concern if exposed to a virulent pathogen such as this putative new biotype of CDV. The Serengeti is surrounded by approximately 30,000 domestic dogs, most of which are not vaccinated against canid pathogens (including CDV), representing a large reservoir for carnivore diseases. The CDV epidemic clearly emphasizes the need for continued disease surveillance to monitor infectious diseases in valuable wildlife resources, and for initiating vaccination programs for domestic animals in contact with wildlife. □

2. Alexander, K. A. & Appel, M. J. G. *J. Wildl. Dis.* **30**, 481–485 (1994).
3. Macdonald, D. W. *Nature* **360**, 633–634 (1992).
4. Packer, C. & Pusey, A. E. *Am. Nat.* **145**, 833–841 (1995).
5. Packer, C. *et al. Reproductive Success* (ed. Clutton-Brock, T. H.) 363–383 (Univ. Chicago Press, 1988).
6. Brown, E. W., Yuhki, N., Packer, C. & O'Brien, S. J. *J. Virol.* **68**, 5953–5968 (1994).
7. Gilbert, D. A., Packer, C., Pusey, A. E., Stephens, J. C. & O'Brien, S. J. *J. Hered.* **82**, 378–386 (1991).
8. Wildt, D. E. *et al. Nature* **329**, 328–331 (1987).
9. Appel, M. J. G. *Virus Infections in Carnivores* (ed. Appel, M. J. G.) 133–160 (Elsevier, Amsterdam, 1987).
10. Appel, M. J. G. *et al. J. vet. diagnost. Invest.* **6**, 277–288 (1994).
11. Orvell, C., Sheshberadaran, H. & Norrby, E. *J. Virol.* **66**, 443–456 (1985).
12. Appel, M. J. G. & Robson, D. S. *Am. J. vet. Res.* **34**, 1459–1463 (1973).
13. Appel, M. J. G., Pearce-Kelling, S. & Summers, B. A. *J. vet. diagnost. Invest.* **4**, 258–263 (1992).
14. Montali, R. J., Bartz, C. R. & Bush, M. *Virus Infections of Carnivores* (ed. Appel, M.) 437–443 (Elsevier, Amsterdam, 1991).
15. Brown, E. W., Miththapala, S. & O'Brien, S. O. *J. Zool. Wildl. Med.* **24**, 357–364 (1993).
16. Carmichael, L. E., Joubert, J. C. & Pollack, R. V. *H. Am. J. vet. Res.* **41**, 784–791 (1980).
17. Hofer, H. & East, M. L. *Symp. zool. Soc. London* **65**, 347–366 (1993).
18. Barrett, T. *et al. Virology* **193**, 1010–1012 (1993).
19. Swafford, D. L. *Phylogenetic Analysis Using Parsimony (PAUP)* (Illinois Natural History Survey, Champaign, IL, 1984).
20. Felsenstein, J. *Phylogeny Inference Package (PHYLP)*, Version 3.3 (University of Washington, Seattle, WA, 1995).

ACKNOWLEDGEMENTS. We thank M. Grob and M. Elgizoli for viral antigens, and M. East and H. Hofer for providing two hyaena carcasses. This work was supported in part by the Messerli Foundation of Zurich, the University of Tennessee, Cornell University, Smithsonian NOAHS Center, the National Science Foundation, National Geographic, and British Airways.

A common molecular basis for three inherited kidney stone diseases

Sarah E. Lloyd*, **Simon H. S. Pearce***,
Simon E. Fisher†, **Klaus Steinmeyer‡**,
Blanche Schwappach‡, **Steven J. Scheinman*§**,
Brian Harding*, **Alessandra Bolino||**,
Marcella Devoto||, **Paul Goodyer¶**, **Susan P. A. Rigden#**,
Oliver Wrong*, **Thomas J. Jentsch‡**,
Ian W. Craig† & **Rajesh V. Thakker*††**

* MRC Molecular Endocrinology Group, Royal Postgraduate Medical School, Hammersmith Hospital, Du Cane Road, London W12 0NN, UK

† Genetics Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

‡ Centre for Molecular Neurobiology Hamburg (ZMNH), Hamburg University, Martinistrasse 52, D-20246 Hamburg, Germany

§ Department of Medicine, State University of New York Health Science Center, Syracuse, New York 13210, USA

|| Laboratorio di Genetica Molecolare, Istituto G. Gaslini, Largo Gaslini 5, 16148 Genova, Italy

¶ L'Hôpital de Montréal pour Enfants, 2300 rue Tupper, A717 Montreal, Quebec H3H1P3, Canada

Department of Paediatric Nephrology, Guy's Hospital, St Thomas's Street, London SE1 9RT, UK

* Department of Nephrology, The Middlesex Hospital, Mortimer Street, London W1N 8AA, UK

KIDNEY stones (nephrolithiasis), which affect 12% of males and 5% of females in the western world, are familial in 45% of patients^{1,2} and are most commonly associated with hypercalciuria¹. Three disorders of hypercalciuric nephrolithiasis (Dent's disease³, X-linked recessive nephrolithiasis (XRN)⁴, and X-linked recessive hypophosphataemic rickets (XLRH)⁵ have been mapped to Xp11.22 (refs 5–7). A microdeletion⁶ in one Dent's disease kindred allowed the identification of a candidate gene, *CLCN5* (refs 8,9) which encodes a putative renal chloride channel. Here we report the investigation of 11 kindreds with these renal tubular disorders for *CLCN5* abnormalities; this identified three nonsense, four missense and two donor splice site

mutations, together with one intragenic deletion and one microdeletion encompassing the entire gene. Heterologous expression of wild-type *CLCN5* in *Xenopus* oocytes yielded outwardly rectifying chloride currents, which were either abolished or markedly reduced by the mutations. The common aetiology for Dent's disease, XRN and XLRH indicates that *CLCN5* may be involved in other renal tubular disorders associated with kidney stones.

Analysis of *CLCN5* reverse transcriptase–polymerase chain reaction (RT–PCR) products encompassing the entire 2238-bp coding sequence⁹ from probands of 11 kindreds with Dent's disease, XRN and XLRH, revealed different *CLCN5* mutations (Table 1 and Fig. 1). Each mutation was confirmed and demonstrated to cosegregate with the disease by using genomic DNA together with the appropriate PCR primers and restriction enzymes, or by sequence-specific oligonucleotide (SSO) probe analysis (Table 1 and Fig. 2). In addition, the absence of these *CLCN5* abnormalities in 110 alleles from 69 (28 males and 41 females) unrelated, normal individuals established that they were not common polymorphisms. *CLCN5* belongs to a family of voltage-gated chloride-channel genes (*CLCN1* to *CLCN5* and *CLCN-Ka* and *Kb*), which encode proteins (CLC-1 to CLC-5, CLC-Ka and CLC-Kb) that have about 12 transmembrane domains¹⁰. These chloride channels are important for the control of membrane excitability, transepithelial transport, and possibly regulation of cell volume¹⁰. However, mutations have been identified previously in only *CLCN1*, which is expressed in muscle¹¹ and is associated with myotonia congenita^{12–14}. Thus, to assess further the functions of CLC-5 and its mutations, we performed heterologous expression studies in *Xenopus* oocytes. Expression of the human wild-type (WT) CLC-5 reproducibly yielded strongly outwardly rectifying, essentially time-independent currents (Fig. 3). Ion substitution experiments indicated that these were carried by anions, with a chloride > iodide conductance sequence (Fig. 3b,c), as is the case with other chloride channels (CLC-0, CLC-1 and CLC-2) of this family^{12,15,16}. However, these human CLC-5 currents, which were indistinguishable from those of rat CLC-5 (ref. 17), differed from the others in being strongly outwardly rectifying and in being observed only at potentials more positive than +10 mV. Although positive potentials of even +40 mV have been observed in apical membranes of some actively transporting epithelia¹⁸, we are not aware of renal cells where these voltages would be reached *in vivo*. CLC channels are known to function as multimeric complexes, which are most likely to be tetramers¹², and it seems possible that CLC-5 forms hetero-oligomers with as yet unknown subunits *in situ* that may render the channels open at a more physiological voltage. Our functional expression of CLC-5 provides a valuable means of investigating this further and in assessing the functional effects of the CLC-5 mutations.

Expression of the four missense and three nonsense mutations (Table 1), and the in-frame deletion of the predicted transmembrane domain D2 (Fig. 1), abolished the CLC-5 currents or reduced them (S244L and S520P) to levels where they were difficult to distinguish from endogenous oocyte currents (Fig. 3d,f). As a control, we also expressed more conservative changes at codons 244 and 520; S244T, S520T, S244A and S520A all elicited chloride currents that were similar to the WT (data not shown). Thus, the mutations found in the hypercalciuric nephrolithiasis pedigrees grossly and specifically affect CLC-5 function, thereby strongly suggesting a causal role in the disease. Dent's disease, which is characterized by low-molecular-weight proteinuria (LMWP), hypercalciuria, nephrocalcinosis, nephrolithiasis, rickets and eventual renal failure^{3,6}, has phenotypic similarities to XRN^{4,7} and XLRH⁵ (Fig. 2). However, there are important differences, as rickets is absent in XRN, and nephrocalcinosis and moderate renal failure are more notable in XLRH. A correlation between these phenotypic differences, the different mutations (Table 1 and Fig. 1) and the resulting abnormal chloride currents (Fig. 3) could not be established. Thus, Dent's disease was found to be associated with the mutations W279X,

††To whom correspondence should be addressed.

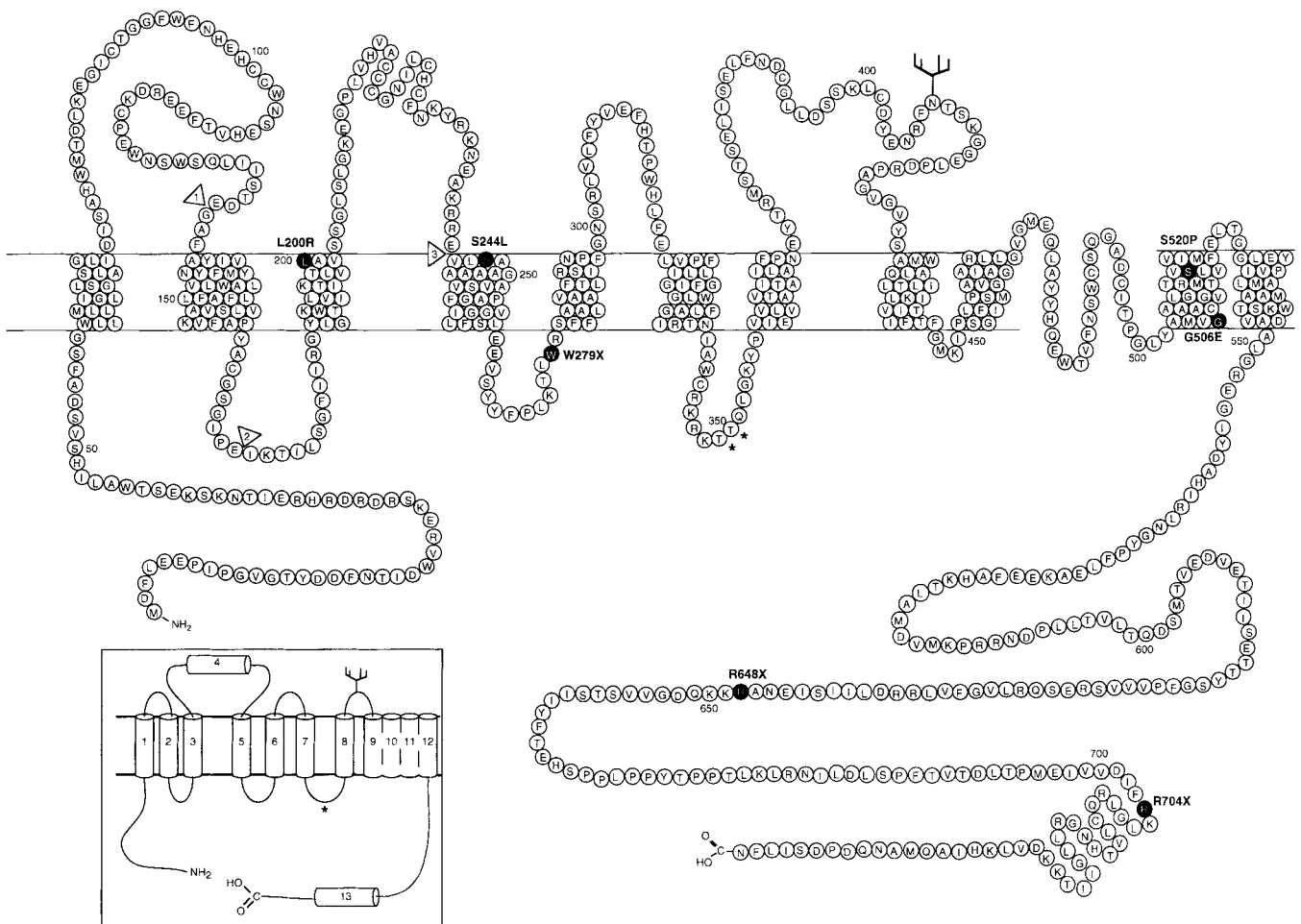


FIG. 1 Schematic representation of a predicted topology of CLC-5, based on the reported DNA sequence⁹, to illustrate the mutations associated with X-linked hypercalcaemic nephrolithiasis. The mutations found in 11 families with X-linked hypercalcaemic nephrolithiasis (Table 1) are shown with the amino acid highlighted in black and the codon and substituted amino acid shown alongside. Arrows 1 and 2 delineate the amino acids predicted to be deleted in families 19/94 and 4/94, and arrows 1–3 delineate those deleted in family 7.1/94 (Table 1). Every 50th amino acid of the 746-amino acid CLC-5 protein and the consensus phosphorylation (asterisk) and glycosylation (branch) sites at codons 349, 350 and 408, respectively, are indicated. The correct topology for the CLC channels is unknown and the predicted topology of the CLC-5 putative transmembrane domains (D1–D13) has been revised⁹ and is based upon a model^{10,15} (inset) that places

D4 extracellularly, and in which the hydrophobic core of the D9–D12 region crosses the membrane 3 or 5 times and contains a hydrophilic region (codons 481–502), the precise location of which remains unknown. An alternative model placing D2 extracellularly has also been proposed^{20,21}, but the donor splice site mutations in families 4/94 and 19/94 that result in the loss of codons 132–172 (arrows 1 and 2) indicate the importance of this segment and provide support for a transmembrane location of D2. The 28 amino acids (codons 700–728) represent a conserved region that has been designated domain D13. The R704X mutation in D13 highlights the importance of this region, the function of which remains to be defined. The nonsense and donor splice site mutations were found to involve the loops and intracellular regions, whereas all the missense mutations occurred within the transmembrane spans.

R648X, L200R, S520P, two microdeletions and two donor splice site mutations; XRN was associated with R704X and G506E; and XLRH was associated with S244L. In addition, the markedly reduced, but not abolished, chloride currents associated with S244L were not specific for XLRH as they were also associated with S520P in Dent's disease (Fig. 3f), and the undetectable chloride currents associated with all the other mutations (Fig. 3) were found in XRN and Dent's disease. As may be expected for a channel protein, all the disease-causing missense mutations were confined to the predicted transmembrane domains (Fig. 1 and Table 1). In addition, the R648X and R704X mutations, which predict a loss of 100 and 42 amino acids from the cytoplasmic carboxy terminus (Fig. 1), delete domain D13, which is conserved in all eukaryotic CLC proteins, including the one of *Saccharomyces cerevisiae*¹⁹. This suggests that D13 and the C-terminal region play an important but as yet unspecified role. The donor splice site mutations (Fig. 2 and Table 1), which are associated with a loss of exon 5, lead to an in-frame deletion of the predicted transmem-

brane domain D2 (Fig. 1). A splice variant of rCLC-K2, deleted for D2, was reported to yield chloride currents indistinguishable from those of WT rCLC-K2 (ref. 20), and it was therefore proposed that D2 does not cross the membrane²¹. However, our findings provide further support for a functional role and transmembrane location for D2.

Our results indicate that CLC-5 is a chloride channel whose functional loss results surprisingly in a renal tubulopathy that is associated with LMWP, hypercalcaemia and nephrolithiasis. The resorption of filtered protein occurs almost exclusively in the proximal tubule, whereas that of calcium occurs in the proximal tubule, the thick ascending limb of Henle's loop, and the distal nephron²². Thus, further investigations of *CLCN5* expression along the human nephron, and of the mechanisms whereby *CLCN5* mutations lead to LMWP and hypercalcaemia, will lead to increased understanding of renal tubular function and pathology. In addition, our findings, which have defined a common molecular pathology of *CLCN5* in three disorders of hereditary

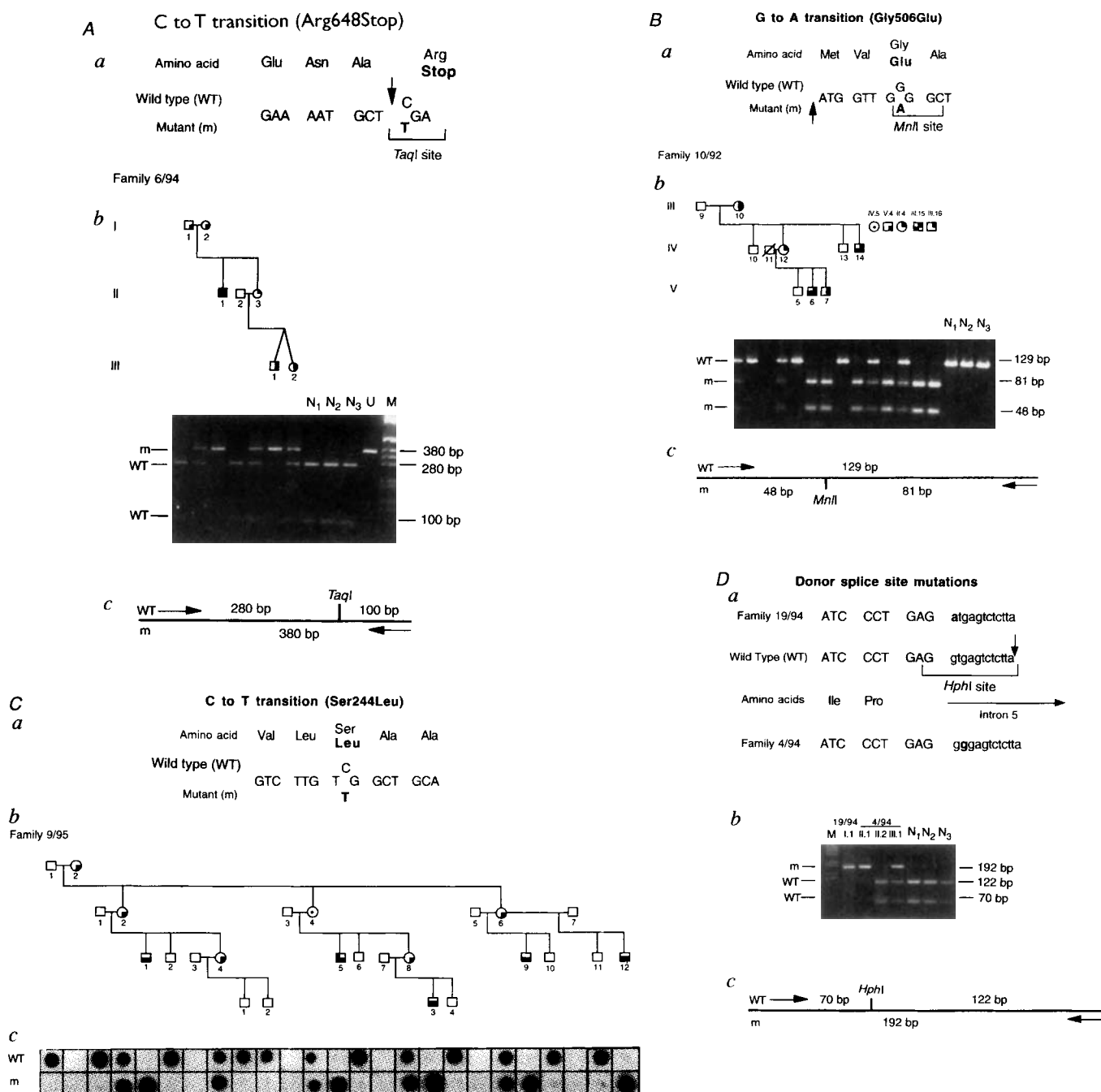


FIG. 2 Detection of CLC-5 mutations by restriction enzyme (A, B, D) and SSO analysis (C). Families with Dent's disease (4/94, 6/94 and 19/94), XRN (10/92) and XLRH (9/95) are shown with LMWP, hypercalciuria, nephrocalcinosis and nephrolithiasis, end-stage renal failure, and carrier female. Only 13 of the 102 members of the six-generation XRN family 10/92 are shown with the individual identification as reported⁶. Family 9/95 (XLRH) has not been assessed for LMWP but III.1, III.5, III.9, III.12 and IV.3 have proteinuria, and III.1 and III.9 have moderate renal failure (creatinine clearance, 60–65 ml min⁻¹ m⁻²)⁵. In family 4/94 (D), II.1 is the affected father, II.2 the unaffected mother, and III.1 the affected daughter. DNA sequence abnormalities, together with their encoded amino acids (a) cosegregation of the mutation with the disease in each family (b), and restriction maps of the wild-type (WT) and mutant (m) sequences (c) are shown. These mutations were not present in 69 unrelated normal individuals (N₁₋₃ shown). Mutant RT-PCR products (192 bp) that differed from WT by 123 bp, which is the size of exon 5, were obtained in families 19/94 and 4/94. This exon skipping resulted from donor splice site mutations and analysis of the exon (upper-case letters) 5–intron (lower-case letters) 5 boundary revealed a g to a transition at position +1 in family 19/94, and t to g transversion at position +2 in family 4/94 (D).

METHODS. RNA was extracted from an Epstein–Barr virus-transformed lymphoblastoid cell line obtained from peripheral blood cells of each proband using methods published previously²³. RT–PCR was performed using pairs of nested CLCN5-specific primers and the following conditions: 25 cycles of denaturation at 94 °C, annealing at 60 °C and extension at 72 °C, each for 30 s. The PCR products were gel purified and the DNA sequences of both strands determined by Taq polymerase cycle sequencing and a semi-automated detection system (ABI-373 sequencer) as described previously²⁴. Genomic DNA, restriction enzyme and SSO studies were performed as described previously²⁵.

Male	Female	
		Low-molecular-weight proteinuria (LMWP)
		Hypercalciuria
		Nephrocalcinosis and nephrolithiasis
		End stage renal failure
		Carrier

TABLE 1 CLCN5 mutations found in families with X-linked hypercalcaemic nephrolithiasis

Family	Codon	Base change	Amino acid change	Restriction enzyme change/SSO	Predicted effect
Nonsense mutations					
7.2/94*	279	TGG → TGA	Trp → Stop	<i>Maelll</i>	Loss of 469 amino acids from D6 to C terminus
6/94*	648	CGA → TGA	Arg → Stop	<i>TaqI</i>	Loss of 100 amino acids from cytoplasmic domain
12/95†	704	CGA → TGA	Arg → Stop	<i>BstI</i>	Loss of 42 amino acids from cytoplasmic domain
Missense mutations					
13/94*	200	CTG → CGG	Leu → Arg	<i>Acil</i>	Disruption of charge distribution within D3
9/95‡	244	TGG → TTG	Ser → Leu	SSO	Disruption of helix in D5
10/92†	506	GGG → GAG	Gly → Glu	<i>MnlI</i>	Disruption of charge distribution within D11
2/92*	520	TCT → CCT	Ser → Pro	<i>MnlI</i>	Disruption of helix in D11
Splice site mutations					
4/94*	Intron 5 132–172 deleted	gt → gg		<i>HphI</i>	Loss of D2
19/94*	Intron 5 132–172 deleted	gt → at		<i>HphI</i>	Loss of D2
Deletions					
7.1/94*	2-kb deletion 132–241 deleted				Loss of D2–D4
12/89*§	515-kb deletion				Absence of protein

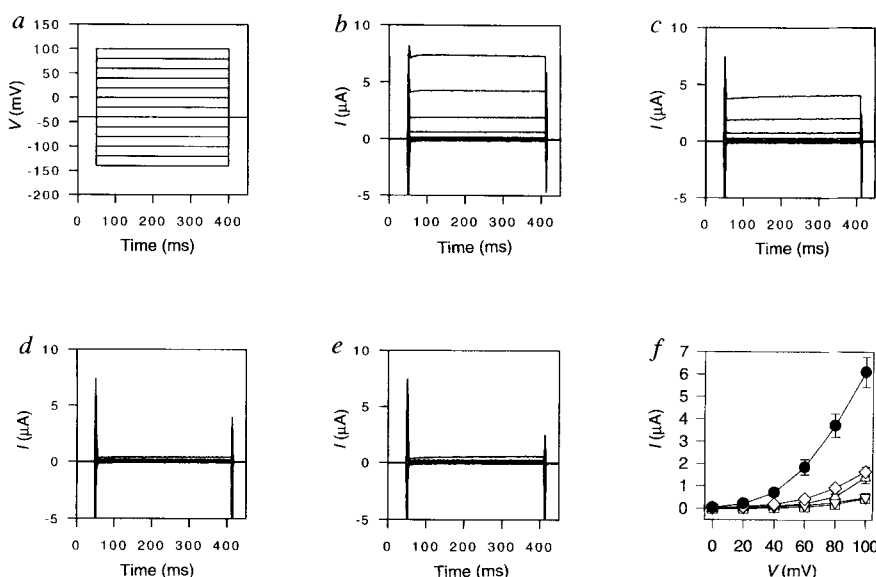
* Eight British families (26 affected, 33 unaffected members) with Dent's disease, † two North American families (29 affected, 75 unaffected members) with XRN, and ‡ one Italian family (9 affected, 10 unaffected members) with XLRH were studied. The clinical details of 7 of the 8 kindreds with Dent's disease, designated families 2/92, 4/94, 6/94, 7.2/94, 12/89, 13/94 and 19/94 have been previously reported³ and referred to as families A, E, D, F, H, C and G, respectively.

§ Previously described^{6,8}.

FIG. 3 Electrophysiological analysis of *Xenopus* oocytes expressing human CLC-5 and its mutants. **a**, The pulse protocol for voltage-clamping, in which oocytes were sequentially clamped from a holding potential of -40 mV to voltages between $+100$ and -140 mV for 350 ms in steps of 20 mV. **b**, The resulting current traces from an oocyte expressing WT CLC-5, measured in ND96 (104 mM chloride), and **c**, in ND96 in which 80 mM Cl^- was replaced by I^- . The resulting strongly outwardly rectifying anion current is reduced by partial replacement of Cl^- by I^- .

d, Current traces from a mutant in which the exon encoding domain D2 (Fig. 1) is deleted in frame (Table 1 and Fig. 2); currents did not differ from water-injected control oocytes (**e**). **f**, Current–voltage relationships showing currents (measured in ND96) of WT CLC-5 (filled circles) and the missense mutants L200R (inverted triangles), S244L (triangles), G506E (squares) and S520P (diamonds). Currents are averaged from several oocytes, and error bars indicate standard error of the mean (s.e.m.). Because CLC-5 is strongly outwardly rectifying, only currents at positive voltages are shown. Although L200R and G506E currents are not different from uninjected control oocytes, mutants S244L and S520P elicit strongly reduced, but detectable, currents. Currents with nonsense mutants (W279X, R648X, R704X) were not different from negative controls (data not shown). Averaged currents at $+80$ mV ($\mu\text{A} \pm \text{s.e.m.}$, n , number of oocytes) were determined to be: WT, 3.71 ± 0.51 , $n = 17$; uninjected control, 0.27 ± 0.03 , $n = 19$; L200R, 0.19 ± 0.03 , $n = 4$; S244L, 0.54 ± 0.12 , $n = 6$; G506E, 0.27 ± 0.02 , $n = 12$; S520P, 0.92 ± 0.11 , $n = 12$; W279X, 0.23 ± 0.02 , $n = 10$; R648X, 0.32 ± 0.08 , $n = 3$; R704X, 0.26 ± 0.03 , $n = 7$; deletion of D2, 0.23 ± 0.03 , $n = 4$.

METHODS. After engineering an *NcoI* site at the initiator methionine, a cDNA encoding CLC-5 was inserted into the *NcoI* site of the expression vector, PTLN, containing *Xenopus* globin untranslated regions²⁶. Mutations



nephrolithiasis, indicate that CLC-5 and possibly other renal chloride channels, may be implicated in other disorders associated with renal stones, which account for up to 1% of all hospital admissions². **Note added in proof:** Two additional genes (*CLCN6* and *CLCN7*) encoding the putative chloride channels CLC-6 and CLC-7 have also recently been cloned²⁸.

Received 13 October; accepted 18 December 1995.

1. Coe, F. L., Parks, J. H. & Asplin, J. R. *New Engl. J. Med.* **327**, 1141–1152 (1992).
2. Smith, L. H. *J. Urol.* **141**, 707–710 (1989).
3. Wong, O. M., Norden, A. G. W. & Feest, T. G. Q. *J. Med.* **87**, 473–493 (1994).
4. Frymoyer, P. A. et al. *New Engl. J. Med.* **325**, 681–686 (1991).
5. Bolino, A. et al. *Eur. J. hum. Genet.* **1**, 269–279 (1993).
6. Pook, M. A. et al. *Hum. molec. Genet.* **2**, 2129–2134 (1993).
7. Scheinman, S. J. et al. *J. clin. Invest.* **91**, 2351–2357 (1993).
8. Fisher, S. E. et al. *Hum. molec. Genet.* **3**, 2053–2059 (1994).
9. Fisher, S. E. et al. *Genomics* **29**, 598–606 (1995).
10. Jentsch, T. J., Günther, W., Pusch, M. & Schwappach, B. *J. Physiol., Lond.* **482**, 19S–25S (1995).
11. Steinmeyer, K., Ortlund, C. & Jentsch, T. J. *Nature* **354**, 301–304 (1991).
12. Steinmeyer, K. et al. *EMBO J.* **13**, 737–743 (1994).
13. Meyer-Kleine, C., Steinmeyer, K., Ricker, K., Jentsch, T. J. & Koch, M. C. *Am. J. hum. Genet.* **57**, 1325–1334 (1995).
14. Koch, M. C. et al. *Science* **257**, 797–800 (1992).
15. Pusch, M., Ludwig, U., Rehfeldt, A. & Jentsch, T. J. *Nature* **373**, 527–531 (1995).
16. Thiemann, A., Gründer, S., Pusch, M. & Jentsch, T. J. *Nature* **356**, 57–60 (1992).
17. Steinmeyer, K., Schwappach, B., Bens, M., Vandewalle, A. & Jentsch, T. J. *J. biol. Chem.* **270**, 31172–31177 (1995).
18. Higgins, J. T., Gebler, B. & Frömter, E. *Pflügers Arch. Eur. J. Physiol.* **371**, 87–97 (1977).
19. Green, J. R., Brown, N. H., DiDomenico, B. J., Kaplan, J. & Eide, D. J. *Molec. gen. Genet.* **241**, 542–553 (1993).
20. Adachi, S. et al. *J. biol. Chem.* **269**, 17677–17683 (1994).
21. Middleton, R. E., Pheasant, D. J. & Miller, C. *Biochemistry* **33**, 13189–13198 (1994).
22. Friedman, P. A. & Gesek, F. A. *Physiol. Rev.* **75**, 429–471 (1995).
23. Sambrook, J., Fritsch, E. F. & Maniatis, T. *A Laboratory manual* 2nd edn. 7.19–7.22 (Cold Spring Harbor Laboratory Press, New York, 1989).
24. Pearce, S. H. S. et al. *J. clin. Invest.* **96**, 2683–2692 (1995).
25. Thakker, R. V. et al. *J. clin. Invest.* **91**, 2815–2821 (1993).
26. Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057–7070 (1984).
27. Jentsch, T. J., Steinmeyer, K. & Schwarz, G. *Nature* **348**, 510–514 (1990).
28. Brandt, S. & Jentsch, T. J. *FEBS Lett.* **377**, 15–20 (1995).

ACKNOWLEDGEMENTS. We thank D. Schlessinger and G. Williams for discussions; B. Farren for collection of samples; A. G. W. Norden for urine analysis; C. Zoccali and G. Romeo for access to the Italian family; and J. Walls for access to one of the British families. S.H.S.P. is an MRC Training Fellow. This work was supported by the Medical Research Council (S.E.L., S.H.S.P., B.H. and R.V.T.), the Human Genome Mapping Project, UK (S.E.F. and I.W.C.), the Deutsche Forschungsgemeinschaft (T.J.J. and K.S.) and the American Heart Association (S.J.S.).

Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli

Jacques Mirenowicz* & Wolfram Schultz†

Institute of Physiology, University of Fribourg, CH-1700 Fribourg, Switzerland

MIDBRAIN dopamine systems are crucially involved in motivational processes underlying the learning and execution of goal-directed behaviour^{1–5}. Dopamine neurons in monkeys are uniformly activated by unpredicted appetitive stimuli such as food and liquid rewards and conditioned, reward-predicting stimuli. By contrast, fully predicted stimuli are ineffective^{6–8}, and the omission of predicted reward depresses their activity⁹. These characteristics follow associative-learning rules^{10,11}, suggesting that dopamine responses report an error in reward prediction¹². Accordingly, neural network models are efficiently trained using a dopamine-like reinforcement signal^{13,14}. However, it is unknown whether the responses to environmental stimuli concern specific motivational attributes or reflect more general stimulus salience^{4,15}. To resolve this, we have compared dopamine impulse responses to motivationally opposing appetitive and aversive stimuli. In contrast to appetitive events, primary and conditioned non-noxious aversive stimuli either failed to activate dopamine neurons or, in cases of close resemblance with appetitive stimuli, induced weaker

responses than appetitive stimuli. Thus, dopamine neurons preferentially report environmental stimuli with appetitive rather than aversive motivational value.

Two monkeys were instrumentally conditioned with visual and auditory stimuli. Appetitive conditioned stimuli eliciting lever pressing for a juice reward alternated randomly with aversive conditioned stimuli inducing hand withdrawal from a resting key in order to avoid a mild air puff to the hand or hypertonic saline to the mouth. Appetitive and aversive stimuli were adjusted to comparable motivational strength by setting juice drop size, air pressure and saline drop size slightly above a threshold below which task performance dropped sharply. Task performance was >90–95% correct. Outside the task, animals were presented with free juice drops but not with air puffs or hypertonic saline, which rapidly disrupted their collaboration.

We used two monkeys to record from dopamine neurons in midbrain catecholamine cell groups A8 (23 neurons dorsal to lateral substantia nigra), A9 (251 neurons in pars compacta of substantia nigra) and A10 (40 neurons in ventral tegmental area); we distinguished these from other midbrain neurons by their distinctive electrophysiological characteristics, which consisted of polyphasic, relatively long discharges occurring at comparatively low frequencies^{6–8,16}. Most dopamine neurons were phasically activated by primary and conditioned appetitive stimuli at short mean latencies of 93–115 ms, thus confirming earlier results^{6–8} (conditioned light: 100 of 128 neurons, 78%; conditioned sound: 120 of 158 neurons, 76%; free juice: 62 of 80 neurons, 78%). In contrast, very few dopamine neurons were activated by aversive stimuli, such as a conditioned sound for air puff (1 of 31 neurons, 3%), light for air puff (8 of 56 neurons, 14%) or fractal picture for saline (4 of 30 neurons, 13%) (Fig. 1). The total of 13 neurons activated by conditioned aversive stimuli belonged to groups A8 (5 of 12 neurons), A9 (7 of 89) and A10 (1 of 16), suggesting a potential mediolateral gradient for the few aversive responses. These few aversive responses failed to result in an average population response (Fig. 3b, d right). The infrequent primary aversive air puff activated only 7 of 51 neurons (14%), all of them being in A9 (Fig. 2). Sixteen of the total 20 neurons with aversive responses also showed appetitive responses. The low responsiveness to aversive stimuli cannot be attributed to movement differences which have been shown not to influence dopamine neurons^{7,17,18}. Conditioned aversive stimuli elicited depressions in 36 of 117 neurons (31%; Fig. 1, bottom), reminiscent of depressions related to the omission of expected reward⁹.

The chronology of experiments demonstrates the way dopamine neurons were preferentially activated by appetitive stimuli. The first animal was extensively conditioned with a small light for appetitive outcome. Later, an adjacent, similar light of a different colour served as the aversive stimulus associated with air-puff avoidance and alternated randomly with the appetitive light, the animal discriminating well between them. Dopamine neurons discriminated quantitatively between these motivationally opposing stimuli, the appetitive light activating the majority of them whereas the similar aversive light activated fewer neurons and at lower magnitudes (Fig. 3a). In order to distinguish between genuinely, albeit lower, aversive responsiveness and confounding generalization to appetitive stimuli, we subsequently conditioned an auditory aversive stimulus previously not associated with appetitive outcome. As already described, only 1 of 31 neurons was activated by the aversive sound (Fig. 3b), suggesting that the aversive activations of Fig. 3a were due to stimulus generalization. In a third step, we tested the general effectiveness of auditory stimuli by reversing stimulus modalities and indeed found most neurons to be activated by the appetitive sound (Fig. 3c). Although behavioural response to the appetitive light had been extinguished, the similar aversive light was still effective, albeit with many fewer neurons and at lower response magnitude. This responsiveness may either suggest a remaining appetitive 'tag' on visual stimuli or a general visual response preference. When testing this, we avoided stimulus generalization by presenting

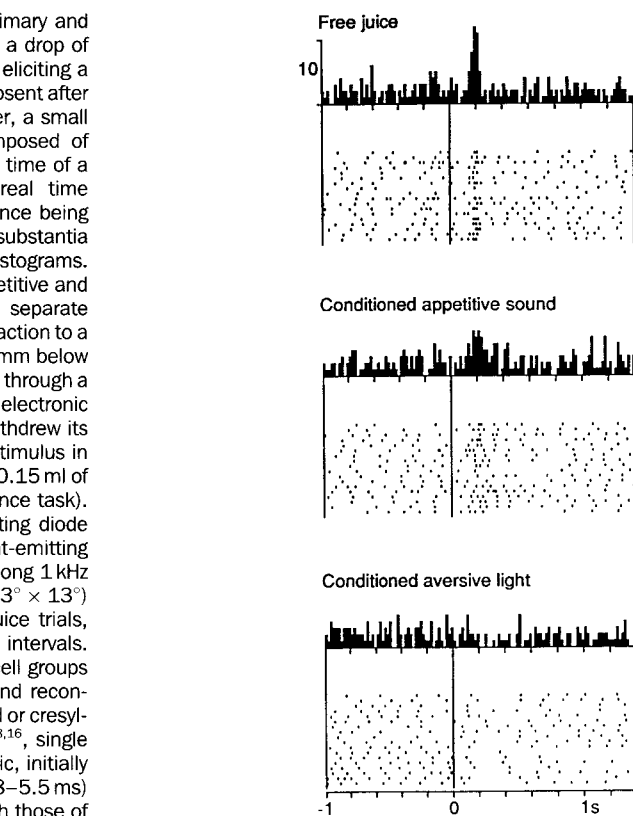
* Present address: Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK.
† To whom correspondence should be addressed.

FIG. 1 Selectivity of activations of one dopamine neuron by primary and conditioned appetitive stimuli. Neuronal activity increased after a drop of free juice outside any task (top) and after a conditioned sound eliciting a reaching movement for juice reward (middle). Activations were absent after a conditioned light in air-puff avoidance trials (bottom). However, a small depression occurred after the stimulus. Histograms are composed of neuronal impulses shown as dots below. Each dot denotes the time of a neuronal impulse, their distances to stimuli representing real time intervals. Each line of dots shows one trial, the original sequence being from top to bottom. The neuron was recorded in group A9 (substantia nigra). Vertical calibration in 10 impulses per bin applies to all histograms. METHODS. All neurons were tested in randomly alternating appetitive and aversive trials. Free-juice trials outside any task were given in separate blocks. In appetitive trials, the animal released a resting key in reaction to a light or sound stimulus and touched a small lever in front at 20 mm below eye level and 27° lateral in order to receive 0.15 ml of apple juice through a spout at the mouth. Liquid arrived at the mouth 55 ms after the electronic pulse activated the liquid valve. In aversive trials, the animal withdrew its hand from the resting key within 1.5 s after a visual or sound stimulus in order to avoid a mild air puff (2–4 bar) to the hand or delivery of 0.15 ml of 1.7 M NaCl through another spout to the mouth (active avoidance task). Conditioned stimuli were (1) a square yellow or red light-emitting diode (11×11 mm) 40 mm above the lever; (2) a green, round light-emitting diode (3 mm diameter) 10 mm above the lever; (3) a 100-ms-long 1 kHz sound 10 mm above the lever; and (4) a coloured fractal image ($13^\circ \times 13^\circ$) on a video screen 260 mm from the animal's eyes. In free-juice trials, animals received the same amount of apple juice at irregular intervals. Recording sites of dopamine neurons randomly sampled from cell groups A8, A9 and A10 were marked with small electrolytic lesions and reconstructed from 40- μ m-thick, tyrosine-hydroxylase-immunoreacted or cresyl-violet-stained coronal brain sections. As previously described^{6–8,16}, single dopamine neurons recorded extracellularly discharged polyphasic, initially negative or positive impulses with relatively long durations (1.8–5.5 ms) and low frequencies (0.5–8.0 imp s⁻¹). Impulses contrasted with those of pars reticulata neurons of substantia nigra (70–90 imp s⁻¹ and <1.1 ms duration) and neighbouring fibres (<0.4 ms duration). Neuronal activations were compared against a 500-ms prestimulus control period by using a Wilcoxon procedure with constant time windows comprising 80% of onset and offset times of statistically significant increases ($P < 0.01$; refs 6–8) (86–325 ms after sound, 76–215 ms after light, 126–270 ms after liquid). Neuronal responsiveness was assessed in terms of (1) numbers

aversive visual stimuli to the second animal before it had ever been conditioned with appetitive visual stimuli. In step four, very few neurons were activated by an aversive light associated with air-puff avoidance (Fig. 3d) and, in step five, by a picture associated with saline avoidance (Fig. 3e), thus arguing against a visual response preference.

Thus, dopamine activations report the appetitive value of primary and conditioned environmental stimuli. Activations hardly occur after non-noxious aversive stimuli in most situations. Only when aversive stimuli appear in close temporal proximity and in random alternation with physically very similar appetitive stimuli, the discrimination loses its all-or-none character and is expressed by a quantitative preference for appetitive stimuli. The activations are broadcast simultaneously through widespread projections to the striatum and frontal cortex. They are particularly effective for increasing dopamine concentration¹⁹ at specific synapses^{20,21}. Apparently complementary functions are carried out by dopamine release detected by voltammetry and dialysis methods, which occurs with a lower time course and in a larger behavioural spectrum including aversive events^{22–25}. Such

FIG. 2 Lack of response of nine dopamine neurons to aversive air puff (right). Most of these neurons were activated by a conditioned appetitive light eliciting a reaching movement for liquid reward (left). Air puffs occurred when the animal failed to release the resting key within 1.5 s after a conditioned auditory stimulus. Rasters to the right display neuronal activity from all trials with air puffs (2–5 trials per neuron), the left showing matching numbers of randomly interspersed appetitive trials from the same neurons at corresponding vertical levels of display.

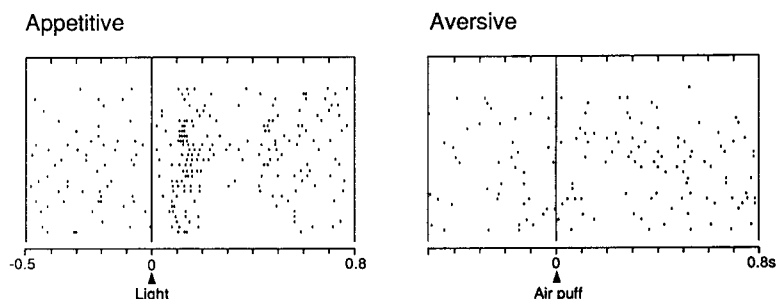


of neurons with significant changes in the time window and (2) magnitude of activation in the time window for every neuron tested, independent of a significant response. All experimental protocols conformed to the Swiss Animal Protection Law and were supervised by the Fribourg Cantonal Veterinary Office.

release appears to be largely due to synaptic overflow²⁶ and possibly extrasynaptic release, and would predominantly affect extrasynaptic receptors²⁷.

The responses to appetitive stimuli and the occasional generalization to physically similar but aversive stimuli apparently reflect an appetitive 'tag' that has been affixed to stimuli through the previous experience of the animal. Stimulus generalization even occurred to a small extent with stimuli whose appetitive associations were behaviourally extinguished (Fig. 3c). Thus dopamine neurons maintain central representations of a maximum range of stimuli that are appetitive or resemble current or previous appetitive stimuli. The responses could be useful in behaviour for exploring new, potentially appetitive stimuli with similar physical features or for testing extinguished appetitive stimuli.

The quick and temporally precise dopamine response to appetitive stimuli may serve two functions. Firstly, the quick activation can be used for the early initiation process of rapid approach reactions. Slower systems with more time for stimulus evaluation could process further stimulus details and provide decisive supplementary information for finalizing behavioural



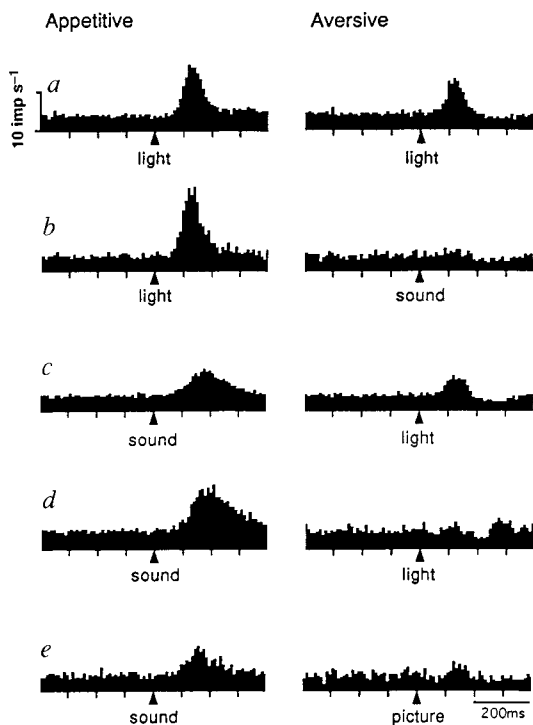


FIG. 3 Chronology of experimentation from a to e shows preferential or exclusive activation of dopamine neurons by appetitive stimuli. Population histograms show neuronal data from the five combinations of randomly alternating appetitive and aversive trials. Data are separated for analysis according to trial type. *a*, Activations in the first monkey elicited by two similar lights serving as appetitive and aversive conditioned stimuli in randomly alternating trials. Of 97 neurons tested, 73 (75%) were activated by an appetitive round green light (mean magnitude, 197% above control) and 63 (65%) by an aversive square yellow light (magnitude 162%; $P < 0.03$; paired *t*-test). *b*, Subsequent separation of sensory modalities revealed an exclusive appetitive activation, thus explaining the aversive activations in *a* with appetitive stimulus generalization. Of 31 neurons, 27 (87%) were activated by the appetitive round green light, but only 1 by the aversive sound. *c*, Subsequent reversal of modalities demonstrated effectiveness of sounds for activating dopamine neurons. Of 72 neurons, 54 (75%) were activated by the appetitive sound (magnitude, 160%), and 23 (32%) by the aversive square red light (magnitude, 100%; $P > 0.01$). *d*, Lack of activation by an aversive square red light in the second monkey which was inexperienced with experimental visual appetitive stimuli. Of 56 neurons, 47 (84%) were activated by the appetitive sound, but only 8 (14%) by the aversive light. Thus, visual aversive activations in *a* and *c* were apparently due to stimulus generalization. *e*, Lack of activation by a coloured fractal picture conditioned as an aversive stimulus in the second monkey. A drop of hypertonic saline served as the primary aversive stimulus, instead of the air puff in *a–d*. Of 30 neurons, 19 (63%) were activated by the appetitive sound, but only 4 (13%) by the aversive picture. The results of *d* and *e* demonstrate the ineffectiveness of aversive stimuli across a considerable range. For each display, histograms from each neuron tested were normalized for trial number, added together, and the resulting sum was divided by the number of neurons. Dopamine neurons were randomly sampled in two monkeys from cell groups A8, A9 and A10, throughout steps *a–e*.

reactions. A candidate is the amygdala with its involvement in reward associations^{28,29} and its highly differentiated responses of longer latencies to reward-associated stimuli³⁰. Secondly, as a temporally precise, global reinforcement signal, the response can exert a temporal selective effect on those synaptic processes that are specifically engaged in the behaviour to be reinforced, thus leading to dopamine-mediated long term changes suggested by the correspondence with learning rules^{12–14}. □

Received 17 July; accepted 23 November 1995.

- Beninger, R. J. & Hahn, B. L. *Science* **220**, 1304–1306 (1983).
- Folger, H. C. & Phillips, A. G. in *Handbook of Physiology—The Nervous System IV*, (ed. Bloom, F. E.) 647–675 (Williams and Wilkins, Baltimore, 1986).

- Wise, R. A. & Rompre, P. P. *Rev. Psychol.* **40**, 191–225 (1989).
- Robbins, T. W. & Everitt, B. J. *Sem. Neurosci.* **4**, 119–128 (1992).
- Robinson, T. E. & Berridge, K. C. *Brain Res. Rev.* **18**, 247–291 (1993).
- Romo, R. & Schultz, W. *J. Neurophysiol.* **63**, 592–606 (1990).
- Ljungberg, T., Apicella, P. & Schultz, W. *J. Neurophysiol.* **67**, 145–163 (1992).
- Mirenowicz, J. & Schultz, W. *J. Neurophysiol.* **72**, 1024–1027 (1994).
- Schultz, W., Apicella, P. & Ljungberg, T. *J. Neurosci.* **13**, 900–913 (1993).
- Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. H. & Prokasy, W. F.) 64–99 (Appleton Century Crofts, New York, 1972).
- Dickinson, A. *Contemporary Animal Learning Theory* (Cambridge Univ. Press, Cambridge, 1980).
- Schultz, W. et al. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J. C., Davis, J. L. & Beiser, D. G.) 233–248 (MIT Press, Cambridge, MA, 1995).
- Sutton, R. S. & Barto, A. G. *Psychol. Rev.* **88**, 135–170 (1981).
- Friston, K. J., Tononi, G., Reeke, G. N., Sporns, O. & Edelman, G. M. *Neuroscience* **59**, 229–243 (1994).
- Schultz, W. *Sem. Neurosci.* **4**, 129–138 (1992).
- Schultz, W. & Romo, R. *J. Neurophysiol.* **57**, 201–217 (1987).
- DeLong, M. R., Crutcher, M. D. & Georgopoulos, A. P. *J. Neurosci.* **3**, 1599–1606 (1983).
- Schultz, W., Ruffieux, A. & Aebischer, P. *Exp. Brain Res. suppl.* **7**, 171–181 (1983).
- Chergui, K., Suaud-Chagny, M. F. & Gonon, F., *Neuroscience* **62**, 641–645 (1994).
- Freund, T. F., Powell, J. F. & Smith, A. D. *Neuroscience* **13**, 1189–1215 (1984).
- Goldman-Rakic, P. S., Leranth, C., Williams, M. S., Mons, N. & Geffard, M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9015–9019 (1989).
- Louilot, A., LeMoal, M. & Simon, H. *Brain Res.* **397**, 395–400 (1986).
- Church, W. H., Justice, J. B. & Neill, D. B. *Brain Res.* **412**, 397–399 (1987).
- Abercrombie, E. D., Keefe, K. A., DiFrischia, D. S. & Zigmond, M. J. *J. Neurochem.* **52**, 1655–1658 (1989).
- Young, A. M., Joseph, M. H. & Gray, J. A. *Neuroscience* **48**, 871–876 (1992).
- Garris, P. A., Ciolkowski, E. L. & Wightman, R. M. *Neuroscience* **59**, 417–427 (1994).
- Yung, K. K. L. et al. *Neuroscience* **65**, 709–730 (1995).
- Gaffan, D. & Harrison, S. J. *Neurosci.* **7**, 2285–2292 (1987).
- Cador, M., Robbins, T. W. & Everitt, B. J. *Neuroscience* **30**, 77–86 (1989).
- Nishijo, H., Ono, T. & Nishino, H. *J. Neurosci.* **8**, 3570–3583 (1988).

ACKNOWLEDGEMENTS. We thank A. Dickinson, J. R. Hollerman, T. W. Robbins and R. A. Wise for comments and suggestions, and B. Aebischer, J. Corpataux, A. Gaillard, A. Pisani, A. Schwarz and F. Tinguely for technical assistance. Supported by the Swiss NSF.

Calcium-dependent interaction of N-type calcium channels with the synaptic core complex

Zu-Hang Sheng, Jens Rettig, Terry Cook & William A. Catterall*

Department of Pharmacology, Box 357280, University of Washington, Seattle, Washington 98195-7280, USA

NEUROTRANSMITTER release is initiated by influx of Ca^{2+} through voltage-gated Ca^{2+} channels^{1,2}, within 200 μs of the action potential arriving at the synaptic terminal³, as the Ca^{2+} concentration increases from 100 nM to $>200 \mu\text{M}$ ⁴. Exocytosis requires high Ca^{2+} concentration, with a threshold of 20–50 μM and half-maximal activation at 190 μM ^{5,6}. The synaptic membrane proteins syntaxin^{7,8}, 25K synaptosome-associated protein (SNAP25)⁹, and vesicle-associated membrane protein (VAMP)/synaptobrevin^{10–12}, are thought to form a synaptic core complex which mediates vesicle docking and membrane fusion^{13–19}. Synaptotagmin may be the low-affinity Ca^{2+} -sensor^{20–24}, but other Ca^{2+} -sensors are involved^{25–27} as residual neurotransmission persists in synaptotagmin-null mutants. Syntaxin binds to N-type Ca^{2+} channels^{7,8,28,29} at a site in the intracellular loop connecting domains II and III³⁰. Here we describe Ca^{2+} -dependent interaction of this site with syntaxin and SNAP25 which has a biphasic dependence on Ca^{2+} , with maximal binding at 20 μM free Ca^{2+} , near the threshold for transmitter release. Ca^{2+} -dependent interaction of Ca^{2+} channels with the synaptic core complex may be important for Ca^{2+} -dependent docking and fusion of synaptic vesicles.

To analyse the Ca^{2+} dependence of association of the cytoplasmic loop connecting domains II and III ($\text{L}_{\text{II-III}}$) of the $\alpha_{1\text{B}}$ -subunit of N-type Ca^{2+} channels with syntaxin 1A, we measured the binding of two fusion proteins tagged with epitopes containing six histidines ($\text{L}_{\text{II-III}}(718–963)$ and $\text{L}_{\text{II-III}}(773–859)$) to recombinant

* To whom correspondence should be addressed.

syntaxin 1A fused to glutathione *S*-transferase (GST) by affinity chromatography on glutathione-Sepharose and immunoblotting. Weak binding was observed when free Ca^{2+} was $1\ \mu\text{M}$ or less (Fig. 1a,b). Binding of $\text{L}_{\text{II-III}}$ fusion proteins was increased for $[\text{Ca}^{2+}]$ from $10\text{--}26\ \mu\text{M}$, and decreased above $37\ \mu\text{M}$ (Fig. 1a,b). Similar results were obtained in buffers with or without Mg^{2+} (Fig. 1, legend), and Mg^{2+} did not affect the binding of $\text{L}_{\text{II-III}}$ (718–963) to syntaxin up to $500\ \mu\text{M}$. Averaged results (Fig. 1c) show that binding of $\text{L}_{\text{II-III}}$ (718–963) increases sharply between 1 and $25\ \mu\text{M}$, then decreases sharply at higher $[\text{Ca}^{2+}]$. The steep concentration dependence implies that two Ca^{2+} ions must

bind to increase binding affinity, and two additional Ca^{2+} ions must bind to decrease affinity.

To examine the Ca^{2+} dependence of syntaxin binding to the intact N-type Ca^{2+} channel, N-type channels were labelled by binding of ω -CTx-GVIA, incubated with GST-syntaxin, and analysed by chromatography on glutathione-Sepharose (Fig. 1d). As with the $\text{L}_{\text{II-III}}$ fusion proteins, weak binding was observed at $0.1\ \mu\text{M}$ free Ca^{2+} , increased binding at $18\ \mu\text{M}$ Ca^{2+} , and reduced binding at $67\ \mu\text{M}$ free Ca^{2+} .

Syntaxin interacts with the synaptic vesicle protein VAMP/synaptobrevin and the synaptic membrane protein SNAP25 to

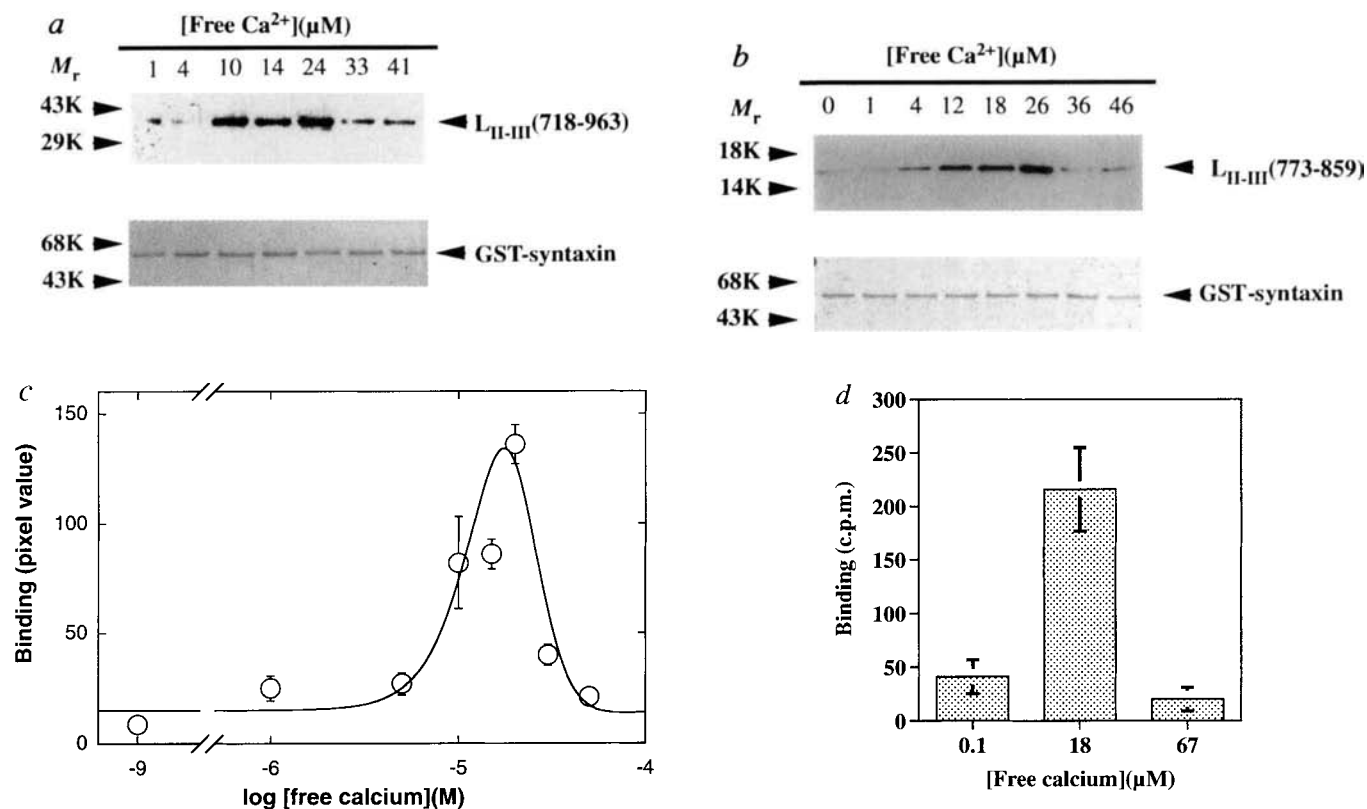


FIG. 1 Ca^{2+} -dependent association of N-type Ca^{2+} channel with syntaxin 1A. Binding of $\text{L}_{\text{II-III}}$ (718–963)(a) and $\text{L}_{\text{II-III}}$ (773–859)(b) to GST-syntaxin was measured in the indicated Ca^{2+} concentrations. GST-syntaxin coupled to glutathione-Sepharose beads was incubated with $\text{L}_{\text{II-III}}$ (718–963) or $\text{L}_{\text{II-III}}$ (773–859), and the bound His-fusion proteins were eluted with glutathione. The larger fusion protein, $\text{L}_{\text{II-III}}$ (718–963), has higher affinity for syntaxin and was used at lower concentrations. Bound His-tag was determined by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis with the T7.Tag antibody, and the presence of equal amounts of GST fusion proteins eluted from the matrix was confirmed by Ponceau S staining of the membrane. Similar results to those illustrated in a and b were obtained with replacement of phosphate buffer by $100\ \text{mM}$ Tris-HCl, pH 7, or in solutions with $5\ \text{mM}$ EGTA as the Ca^{2+} buffer, or in the presence of $1\ \text{mM}$ Mg^{2+} in $5\ \text{mM}$ EGTA as Ca^{2+} buffer (data not shown). GST alone exhibited no Ca^{2+} -dependent binding (data not shown). c, Quantitative analysis of binding data of $\text{L}_{\text{II-III}}$ (718–963) to GST-syntaxin as a function of $[\text{Ca}^{2+}]$. ECL signal intensities were quantified with ImageQuant software (Molecular Dynamics), and the total pixel value of each band was normalized based on the ECL signal intensity of eluted GST-syntaxin on the same blots, and plotted versus free $[\text{Ca}^{2+}]$. Data from five experiments were averaged, and error bars represent s.e.m. d, Ca^{2+} -dependent interaction of ^{125}I - ω -CTx-GVIA-labelled N-type Ca^{2+} channels and GST-syntaxin. Equal amounts (c.p.m.) of N-type Ca^{2+} channels labelled with ^{125}I - ω -CTx-GVIA were incubated with affinity matrices containing $150\ \text{nM}$ GST-syntaxin for $3\ \text{h}$ in HEDTA- Ca^{2+} buffer containing 0.1 , 18 or $67\ \mu\text{M}$ Ca^{2+} as indicated. The beads were washed three times with HEDTA- Ca^{2+} buffer, and the

amount of bound N-type Ca^{2+} channels was assessed by direct counting. The data from three binding experiments were averaged; error bars represent s.d.

METHODS. GST fusion proteins ($150\ \text{nM}$) were bound to glutathione-Sepharose beads (Pharmacia) in buffer A (in mM) ($140\ \text{NaCl}$, $2.7\ \text{KCl}$, $10.1\ \text{Na}_2\text{HPO}_4$, $1.8\ \text{KH}_2\text{PO}_4$, pH 7.3) containing 0.1% Triton X-100, pepstatin, aprotinin, leupeptin (each at $4\ \mu\text{g}\ \text{ml}^{-1}$), $0.4\ \mu\text{M}$ phenylmethane-sulphonyl fluoride, incubated at 4°C for $1\ \text{h}$ with constant agitation, and washed with buffer A to remove uncoupled proteins. The amount of each fusion protein in the binding assays was estimated with a standard curve relating the intensity of the immunoblotting signal to the amount of a standard fusion protein applied. Glutathione-Sepharose beads coupled with similar amounts of GST fusion protein were added to the lysates containing $100\ \text{nM}$ $\text{L}_{\text{II-III}}$ (718–963), or $6\ \mu\text{M}$ $\text{L}_{\text{II-III}}$ (773–859) fusion proteins and incubated with gentle mixing for $3\ \text{h}$ at 4°C . The binding reactions were performed in a Ca^{2+} -buffering system ($5\ \text{mM}$ HEDTA- Ca^{2+} , $50\ \text{mM}$ HEPES, pH 7, in buffer A) containing Ca^{2+} concentrations estimated using the Max Chelator software (version 6.63). The beads were washed three times with the $5\ \text{mM}$ HEDTA- Ca^{2+} buffer and the bound proteins eluted in $15\ \text{mM}$ reduced glutathione and $50\ \text{mM}$ Tris-HCl, pH 8. Eluates were separated from the beads by centrifugation at $10,000\ \text{g}$ for $1\ \text{min}$ and processed for SDS-PAGE and immunoblotting with the T7.Tag antibody (Novagen). N-type Ca^{2+} channels were solubilized with digitonin, partially purified by chromatography on WGA-Sepharose, and labelled with $500\ \text{fmol}$ ^{125}I -Tyr-22- ω -CTx GVIA³⁰.

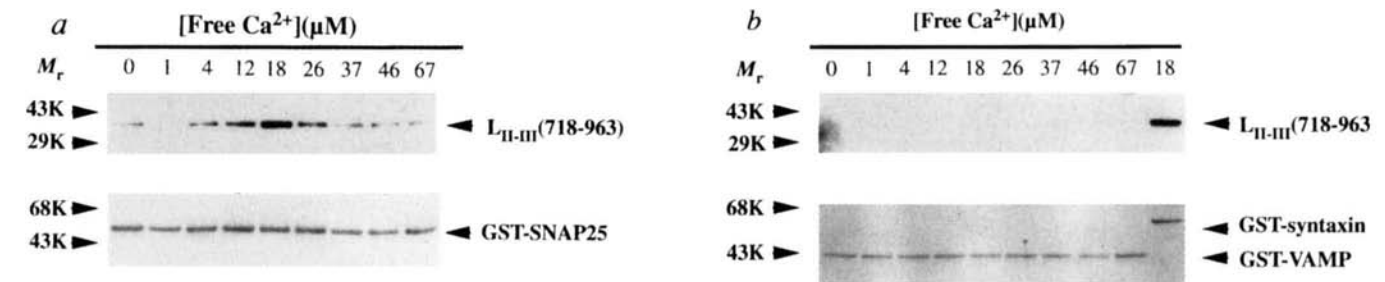


FIG. 2 Ca^{2+} -dependent interaction of N-type Ca^{2+} channel loop with SNAP25. Binding analysis of N-type Ca^{2+} channel fusion protein His-L_{II-III}(718-963) to GST-SNAP25 (a) and GST-VAMP/syntaxin (b) was performed as for Fig. 1. The presence of bound His-tagged N-type

Ca^{2+} channel fusion proteins was determined by western blot analysis with T7.Tag antibody to His-tag sequence, and the presence of equal amounts of GST fusion proteins attached and eluted from matrix was confirmed by Ponceau S staining.

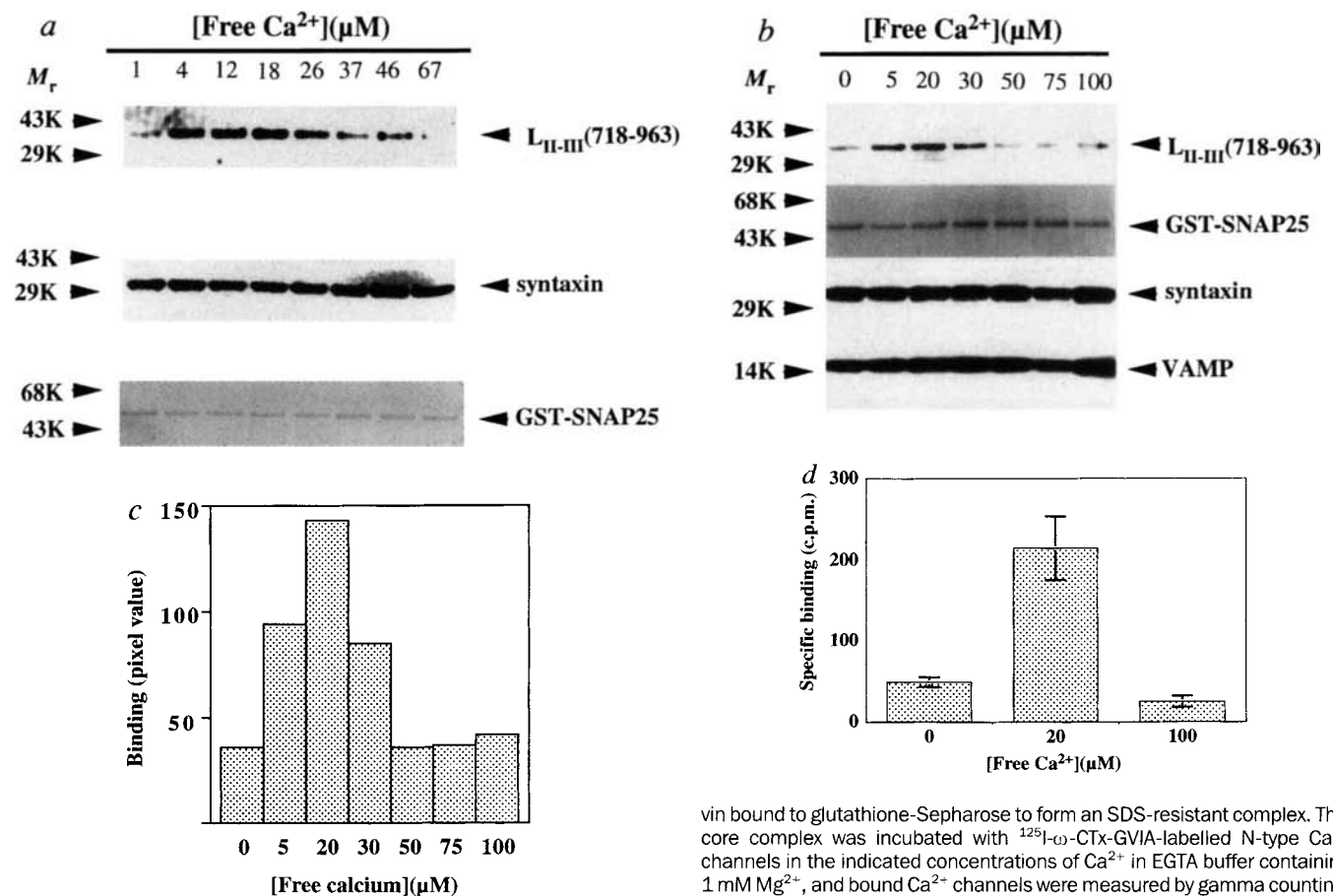


FIG. 3 Ca^{2+} -dependent interaction of N-type Ca^{2+} channel with SNARE protein complexes. a, GST-SNAP25 was bound to glutathione-Sepharose beads and incubated with purified recombinant syntaxin for 2 h in buffer B (100 mM Tris-HCl, pH 7, 140 mM NaCl, 0.1% Triton X-100) to form a recombinant protein dimeric complex. L_{II-III}(718-963) was then incubated with the immobilized complex in increasing [Ca^{2+}]. Binding was measured as in Fig. 1. Top, L_{II-III}(718-963) immunoblotted with anti-T7.Tag antibody; middle, syntaxin immunoblotted with anti-syntaxin antibody; bottom, GST-SNAP25 stained with Ponceau S. b, GST-SNAP25 was bound to glutathione-Sepharose beads and incubated with purified recombinant syntaxin and VAMP/syntaxin for 2 h in buffer B to form a trimeric core complex. Interaction of L_{II-III}(718-963) with the core complex was measured in 5 mM EGTA Ca^{2+} buffer in the presence of 1 mM free Mg^{2+} . The syntaxin-SNAP25-VAMP trimeric complex was visualized with the indicated antibodies. c, Averaged results for experiments like the one in b. d, Calcium-dependent interaction of ^{125}I - ω -CTx-GVIA-labelled N-type Ca^{2+} channels with a native complex of syntaxin-SNAP25-VAMP/syntaxin. N-type Ca^{2+} channels were purified from rat-brain synaptic membranes and labelled with ^{125}I - ω -CTx-GVIA³⁰. Syntaxin and SNAP25 were solubilized from rat-brain synaptic membranes and incubated with GST-VAMP/syntaxin-

vin bound to glutathione-Sepharose to form an SDS-resistant complex. This core complex was incubated with ^{125}I - ω -CTx-GVIA-labelled N-type Ca^{2+} channels in the indicated concentrations of Ca^{2+} in EGTA buffer containing 1 mM Mg^{2+} , and bound Ca^{2+} channels were measured by gamma counting. The data from six binding experiments were averaged; error bars represent s.e.m. Addition of a 200-fold excess concentration of unlabelled ω -CTx-GVIA prevented specific binding and reduced bound ω -CTx-GVIA to the background level observed with GST-VAMP/SNAP25 bound to glutathione-Sepharose alone (data not shown).

METHODS. In a and b, syntaxin and SNAP25 were expressed as GST fusion proteins, purified by affinity chromatography on glutathione-Sepharose, and cleaved from the GST tag with thrombin. Synaptosomes from rat brain were solubilized in buffer C (in mM) (10 HEPES, 1 EGTA, 0.15 M NaCl, 2 MgCl_2 , 1% Triton X-100), and insoluble proteins were removed by centrifugation. When incubated with GST-VAMP/syntaxin affinity matrix for 3 h, an SDS-resistant complex including syntaxin and SNAP25 was observed (data not shown). N-type Ca^{2+} channels are much less abundant and are not detected in the SDS-resistant complex (data not shown). Equal amounts (c.p.m.) of N-type Ca^{2+} channels from preparations purified by chromatography on WGA-Sepharose and labelled with ^{125}I - ω -CTx-GVIA³⁰ were incubated with GST-VAMP/syntaxin affinity matrices containing the native core complex for 12 h in buffer B with 5 mM EGTA, 1 mM free Mg^{2+} , and 0, 20 or 100 μM of free Ca^{2+} as indicated. The beads were washed twice with the EGTA- Ca^{2+} buffer containing 1 mM Mg^{2+} , and the amount of bound N-type Ca^{2+} channels was assessed by direct counting.

form a high-affinity core complex¹³⁻¹⁹. Like syntaxin, SNAP25 associates with L_{II-III}(718-963) in a Ca²⁺-dependent manner with optimal binding at 18 μ M Ca²⁺ (Fig. 2a), whereas VAMP/synaptobrevin does not (Fig. 2b).

To test the interaction with the core complex, we formed syntaxin-SNAP25 (Fig. 3a) and syntaxin-SNAP25-VAMP/synaptobrevin (Fig. 3b) complexes immobilized on glutathione-Sepharose, and incubated them with L_{II-III}(718-963). Approximately equimolar ratios of these proteins were associated, as assessed by Ponceau S staining. L_{II-III}(718-963) bound the syntaxin-SNAP25 dimeric complex in a Ca²⁺-dependent manner, with strongest binding in 4–18 μ M Ca²⁺ (Fig. 3a). Similar Ca²⁺ dependence was observed for binding to the trimeric core complex of syntaxin, SNAP25 and VAMP/synaptobrevin measured in the presence of 1 mM Mg²⁺ (Fig. 3b). We also measured the effect of [Ca²⁺] on binding of the native N-type Ca²⁺ channel specifically labelled with ω -CTx-GVIA to a complex of native syntaxin and SNAP25 with GST-VAMP/synaptobrevin bound to glutathione-Sepharose in the presence of Mg²⁺ (Fig. 3c). Basal binding is observed in the absence of Ca²⁺, maximal binding at 20 μ M Ca²⁺, and reduced binding at 100 μ M Ca²⁺. Thus Ca²⁺ may regulate the interaction of N-type Ca²⁺ channels with the synaptic-vesicle docking/fusion core complex.

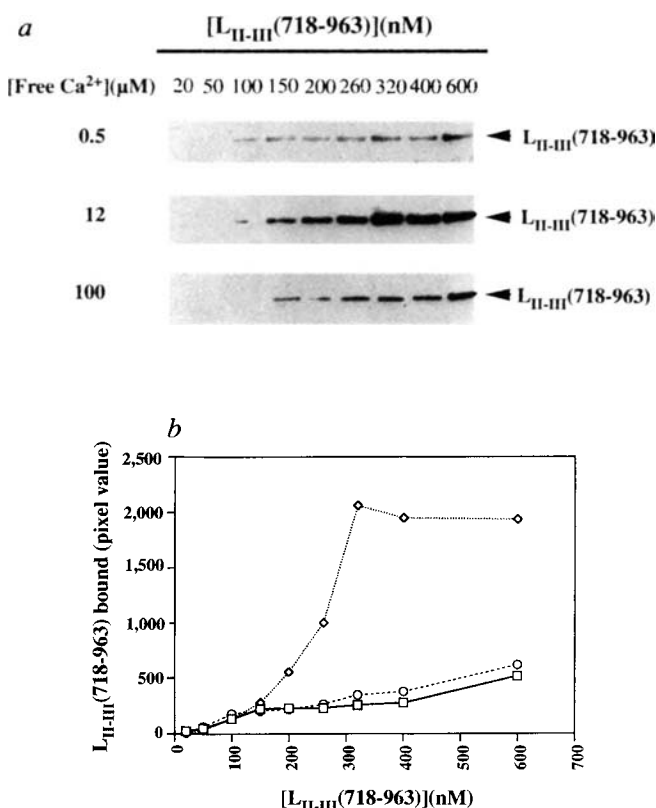


FIG. 4 Binding affinities of N-type Ca²⁺ channel loop L_{II-III} for syntaxin in different Ca²⁺ concentrations. Immunoblots (a) and quantitative studies (b) of the binding affinities of L_{II-III}(718-963) for GST-syntaxin as a function of the Ca²⁺ concentration. His-tagged L_{II-III}(718-963) at the indicated concentrations was bound to immobilized GST-syntaxin (330 nM) in the presence of 0.5 (squares), 12 (diamonds) or 100 μ M of free Ca²⁺ stabilized by 5 mM HEDTA. Binding and immunoblots were as in Fig. 1. The ECL signal intensities were quantitated with ImageQuant software (Molecular Dynamics) and the total pixel value of each band (y axis) is plotted versus the concentration of His-tagged L_{II-III}(718-963). The binding curve appears sigmoid, but this may reflect the non-equilibrium binding assay used in these experiments or the steep, biphasic effect of Ca²⁺, and should not be interpreted as cooperative binding of the Ca²⁺ channel fusion protein from these results.

To estimate the binding affinity of L_{II-III}(718-963) for syntaxin, increasing concentrations of L_{II-III}(718-963) were added to 330 nM GST-syntaxin in 0.5, 12 and 100 μ M Ca²⁺. Binding of L_{II-III}(718-963) in 12 μ M Ca²⁺ increases progressively from 50 to 320 nM and then reaches saturation (Fig. 4a,b). In contrast, binding of L_{II-III}(718-963) in either 0.5 or 100 μ M Ca²⁺ increases progressively, and does not reach saturation up to 600 nM (Fig. 4a,b). These results are consistent with a large (>10-fold), biphasic change in affinity for binding of L_{II-III}(718-963) to syntaxin as a function of [Ca²⁺]. The biphasic dependence of binding affinity on Ca²⁺ suggests Ca²⁺ binding to at least two classes of sites with different affinities and opposite effects on binding of syntaxin and SNAP25 to the Ca²⁺ channel.

Syntaxin is present as a complex with SNAP25 in the pre-synaptic membrane¹³⁻¹⁹. During vesicle docking, syntaxin-SNAP25 dimers form a stable complex with VAMP/synaptobrevin. When neurotransmitter release is initiated, this core complex binds the fusion proteins α -SNAP and NSF, which hydrolyse ATP as part of the driving force for membrane fusion. Ca²⁺ binding to synaptotagmin is part of the signal that initiates rapid exocytosis²⁰⁻²⁴. Our results add two new elements to this evolving model. First, we find that syntaxin, SNAP25 and the synaptic core complex interact with the N-type Ca²⁺ channel through L_{II-III}, supporting the idea that synaptic vesicles are docked near pre-synaptic Ca²⁺ channels through this interaction. Second, we show that this interaction is dependent on changes in [Ca²⁺] near the threshold for initiation of transmitter release. We suggest that docked synaptic vesicles form a low-affinity complex with the Ca²⁺ channel through binding to syntaxin and SNAP25 at resting [Ca²⁺]. Ca²⁺ influx greatly increases the affinity of this complex, so the binding energy of Ca²⁺ to the docked complex may contribute to the energetic driving force for the early priming steps of the fusion process. Finally, as the free [Ca²⁺] reaches the threshold for release (20–50 μ M), the binding affinity to the Ca²⁺ channel is reduced, and syntaxin and SNAP25 dissociate from the presynaptic Ca²⁺ channel to allow membrane fusion to occur, perhaps regulated by Ca²⁺ binding to synaptotagmin and/or vesicle phospholipids in the range 100–200 μ M. The Ca²⁺-dependent interaction of the core complex with the Ca²⁺ channel may contribute to the residual Ca²⁺ dependence of uncoordinated transmitter release in synaptotagmin-deficient mutants²⁵⁻²⁷. □

Received 4 October; accepted 24 November 1995.

- Smith, S. J. & Augustine, G. J. *Trends Neurosci.* **11**, 458–464 (1988).
- Robitaille, R., Adler, E. M. & Charlton, M. P. *Neuron* **5**, 773–779 (1990).
- Barrett, E. F. & Stevens, C. F. *J. Physiol., Lond.* **227**, 691–708 (1972).
- Linás, R., Sugimori, M. & Silver, R. B. *Science* **256**, 677–679 (1992).
- Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. *Nature* **371**, 513–515 (1994).
- von Gersdorff, H. & Matthews, G. *Nature* **367**, 735–739 (1994).
- Bennett, M. K., Calakos, N. & Scheller, R. H. *Science* **257**, 255–259 (1992).
- Yoshida, A. et al. *J. Biol. Chem.* **267**, 24925–24928 (1992).
- Oyler, G. A. et al. *J. Cell Biol.* **109**, 3039–3052 (1989).
- Trimble, D. M., Cowan, D. M. & Scheller, R. H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4538–4542 (1988).
- Elferink, L. A., Trimble, W. S. & Scheller, R. H. *J. Biol. Chem.* **264**, 11061–11064 (1989).
- Südhof, T. C. & Jahn, R. *Neuron* **6**, 665–677 (1991).
- Söllner, T. et al. *Nature* **363**, 318–324 (1993).
- Söllner, T., Bennett, M. K., Whiteheart, S. W., Scheller, R. H. & Rothman, J. E. *Cell* **75**, 409–418 (1993).
- Söllner, T. & Rothman, J. E. *Trends Neurosci.* **17**, 344–348 (1994).
- Pevsner, J. et al. *Neuron* **13**, 353–361 (1994).
- Calakos, N., Bennett, M. K., Peterson, K. & Scheller, R. H. *Science* **263**, 1146–1149 (1994).
- Hayashi, T. et al. *EMBO J.* **13**, 5051–5061 (1994).
- DeBello, W. M. et al. *Nature* **373**, 626–630 (1995).
- Perin, M. S., Fried, V. A., Mignery, G. A., Jahn, R. & Südhof, T. C. *Nature* **345**, 260–263 (1990).
- Brose, N., Petrenko, A. G., Südhof, T. C. & Jahn, R. *Science* **256**, 1021–1025 (1992).
- Davletov, B. A. & Südhof, T. C. *J. Biol. Chem.* **268**, 26386–26390 (1993).
- Chapman, E. R. & Jahn, R. *J. Biol. Chem.* **269**, 5735–5741 (1994).
- Geppert, M. et al. *Cell* **79**, 717–727 (1994).
- DiAntonio, A. K., Parfitt, D. & Schwarz, T. L. *Cell* **72**, 1281–1290 (1993).
- Nonet, M., Grundahl, K., Meyer, B. J. & Band, J. B. *Cell* **73**, 1291–1305 (1993).
- Littleton, J. T., Stern, M., Schulze, K., Perin, M. & Beller, H. J. *Cell* **74**, 1125–1134 (1993).
- O'Connor, V. M., Shamotienko, O., Grishin, E. & Betz, H. *FEBS Lett.* **326**, 255–261 (1993).
- Lévesque, C. et al. *J. Biol. Chem.* **269**, 6309–6312 (1994).
- Sheng, Z. H., Rettig, J., Takahashi, M. & Catterall, W. A. *Neuron* **13**, 1301–1313 (1994).

ACKNOWLEDGEMENTS. We thank M. C. Wilson for SNAP25 complementary DNA; T. P. Snutch for cDNA encoding the α_1 subunit of N-type Ca²⁺ channels; A. Yoshida, Y. Kasai and M. Takahashi for syntaxin cDNA; B. Peterson for help with data analysis; and S. Painter and C. Baker for technical assistance. This work was supported by the NIH and Deutsche Forschungsgemeinschaft.

Two mechanisms of quantized calcium release in skeletal muscle

M. G. Klein*, H. Cheng†‡, L. F. Santana†, Y.-H. Jiang*, W. J. Lederer† & M. F. Schneider*

Departments of Biological Chemistry* and Physiology†, University of Maryland School of Medicine, 660 W. Redwood Street, Baltimore, Maryland 21201, USA

SKELETAL muscle uses voltage sensors in the transverse tubular membrane¹⁻³ that are linked by protein-protein interactions⁴⁻⁶ to intracellular ryanodine receptors⁷⁻¹⁰, which gate the release of calcium from the sarcoplasmic reticulum. Here we show, by using voltage-clamped single fibres and confocal imaging, that stochastic calcium-release events, visualized as Ca^{2+} sparks, occur in skeletal muscle and originate at the triad. Unitary triadic Ca^{2+} -release events are initiated by the voltage sensor in a steeply voltage-dependent manner, or occur spontaneously by a mechanism independent of the voltage sensor. Large-amplitude events also occur during depolarization and consist of two or more unitary events. We propose a 'dual-control' model for discrete Ca^{2+} release events from the sarcoplasmic reticulum that unifies diverse observations about Ca^{2+} -signalling in frog skeletal muscle, and that may be applicable to other excitable cells.

Muscle fibres containing the calcium indicator fluo-3 (ref. 11) exhibited highly non-uniform increases in calcium concentration, $[\text{Ca}^{2+}]$, during small depolarizing pulses (Fig. 1). Fluorescence confocal line-scan images¹²⁻¹⁴ showed discrete foci of locally elevated $[\text{Ca}^{2+}]$ occurring at different times and positions during pulses from -90 to -70 mV (Fig. 1a, b, top). The local $[\text{Ca}^{2+}]$ was distinctly different from the average measurement (top trace below images), and the spatio-temporal pattern of locally elevated $[\text{Ca}^{2+}]$ was variable for repeat depolarizations of the same amplitude (Fig. 1a, b). These observations suggest that discrete stochastic Ca^{2+} -release events occur during depolarization.

The fluo-3 fluorescence before depolarization exhibited a very small ($3.5 \pm 1.1\%$; $N = 3$) but distinct periodic variation in intensity along the scan line (Fig. 2a), possibly resulting from fluo-3 binding to sarcomeric proteins¹⁵ or differences in dye-accessible volume along the sarcomere. A fluorescent extracellular marker¹⁶ located the transverse tubules at the centre of the dark band of the fluo-3 fluorescence pattern of the resting fibre (Fig. 2a). The dark band thus marks the triad location at the Z-line, where activation occurs in these frog skeletal fibres¹⁷, and allowed recording of single triadic Ca^{2+} -release events. We calculated the local change in $[\text{Ca}^{2+}]$, and these $\Delta[\text{Ca}^{2+}]$ records (Fig. 1a, b, bottom) for each of four individual triads (arrowheads (a and b)). The pulses to -70 mV exhibited brief yet clearly discernible elevations in local $[\text{Ca}^{2+}]$. These correspond to triadic Ca^{2+} -release events during which sarcomeric $\Delta[\text{Ca}^{2+}]$ rose rapidly to a peak amplitude several times as large as the mean $\Delta[\text{Ca}^{2+}]$ over the scan line. The individual triadic Ca^{2+} -release events exhibited a range of latencies, with discrete events appearing throughout, but not after, the pulse. A depolarization to -60 mV (Fig. 1c) produced a considerably larger change in $[\text{Ca}^{2+}]$, consistent with the fusion of many discrete release events occurring at higher frequency.

Autocorrelation analysis of release events at different triads indicated that they were uncorrelated and thus strictly independent. The number of triads exhibiting a given number of release events during pulses to -70 mV followed a Poisson distribution (Fig. 2b), consistent with all triads exhibiting the same increase in probability of random release events during depolarization. This

supports the idea that the observed Ca^{2+} -release events in the sarcoplasmic reticulum (SR) were stochastic. The occurrence of Ca^{2+} -release events was steeply voltage dependent, as found previously from whole-fibre $[\text{Ca}^{2+}]$ measurements in this voltage range¹⁸, with the frequency of events increasing exponentially for 4.1 mV of depolarization over the voltage range -80 to -68 mV in

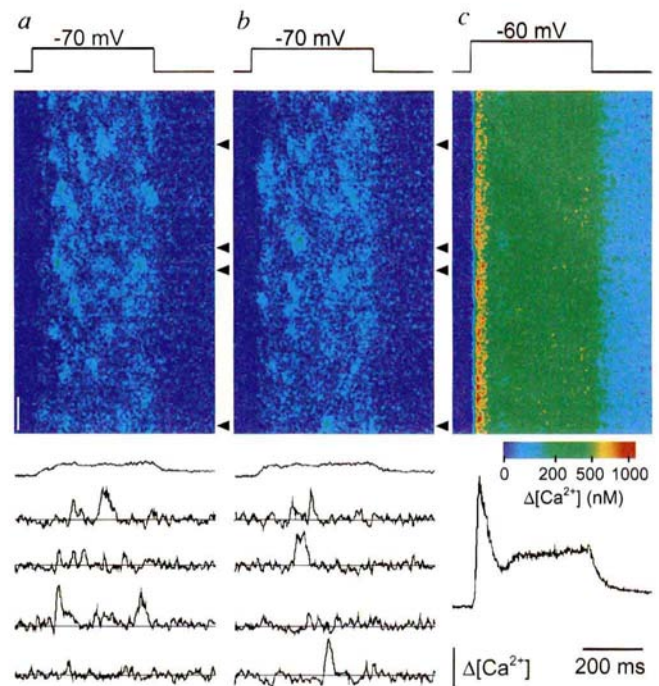
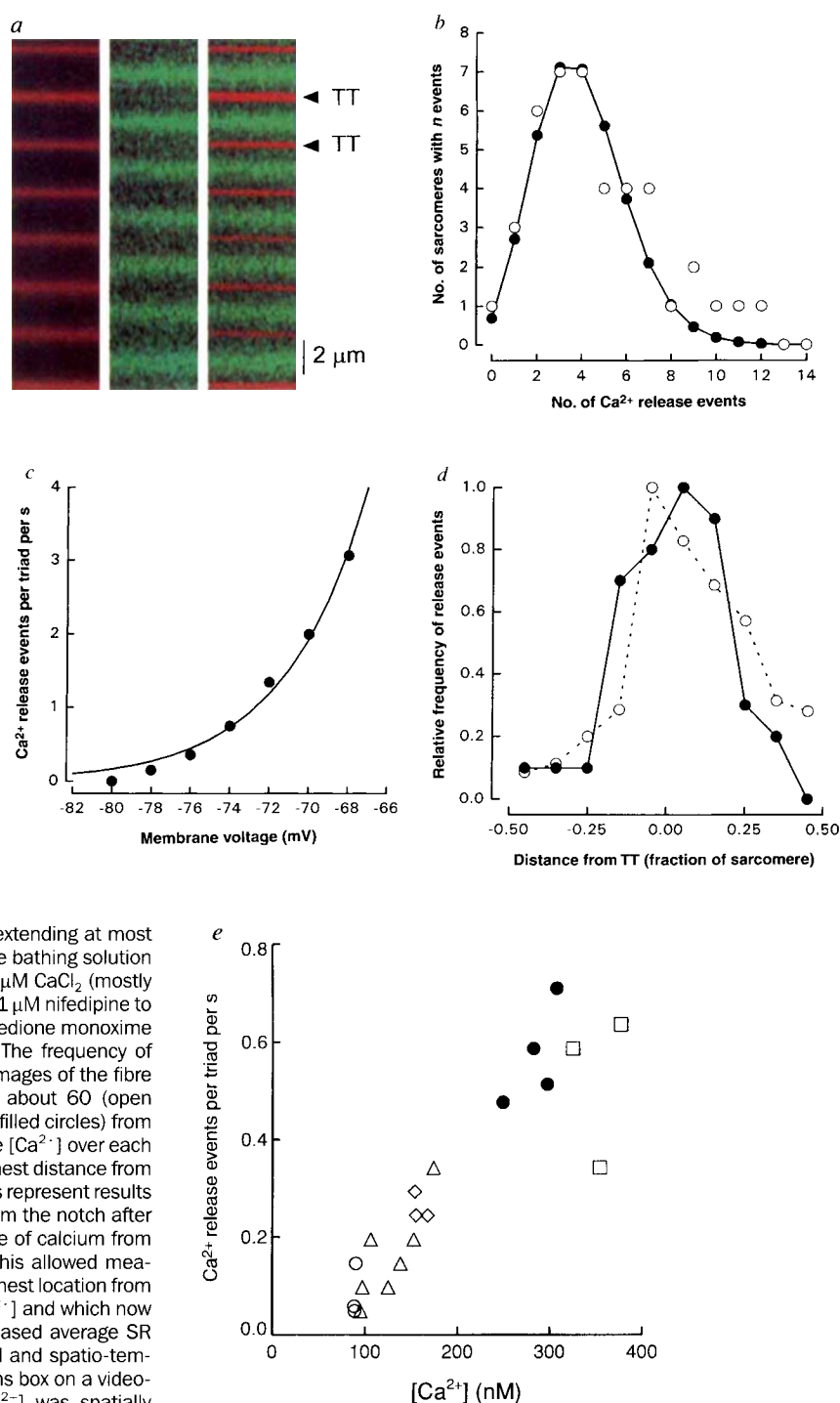


FIG. 1 Local calcium release in skeletal muscle during excitation-contraction coupling. Cut skeletal muscle fibres (ileofibularis) from the frog (*Rana pipiens*) were mounted in an inverted²⁶ double Vaseline gap²⁷ chamber. The fibres were bathed by external solution in the centre pool, and the fibre ends were permeabilized by 0.5-min exposure to 0.01% saponin²⁸ before being bathed by an internal solution. Fibres were loaded with the calcium indicator fluo-3 (ref. 11) by diffusion from the end pools, voltage-clamped to -90 mV, and observed using an inverted confocal microscope system ($0.4 \mu\text{m}$ resolution in x, y and $0.8 \mu\text{m}$ resolution in z)¹²⁻¹⁴. Line-scan fluorescence images were obtained along a line parallel to the long axis of the fibre. Successive vertical lines in each image reveal the time course (left to right) at 2-ms resolution of the fluorescence signal along the scanned line (top to bottom, $0.18 \mu\text{m}$ per pixel) before, during and after 400 ms depolarizations (top) from -90 to -70 , to -70 and to -60 mV (a-c). The traces below the images are the time courses of the spatially averaged mean $\Delta[\text{Ca}^{2+}]$ (a-c) and the local $\Delta[\text{Ca}^{2+}]$ in each of the four single sarcomeres identified by the arrowheads (a and b). The mean $\Delta[\text{Ca}^{2+}]$ was subtracted from each of the individual traces in a and b, with zero given by straight line. The mean $\Delta[\text{Ca}^{2+}]$ in c shows a decline from the early peak, consistent with $[\text{Ca}^{2+}]$ -dependent partial inactivation of release^{29,30} at the level of $[\text{Ca}^{2+}]$ attained during this pulse. The continued occurrence of discrete Ca^{2+} -release events due to the non-inactivating component of release during the pulse in c was indicated by an increased variance/mean ratio of the fluorescence signal, which would be constant if the fluctuation in the fluorescence signal were solely due to photon noise. However, the variance/mean ratio after the pulses was not significantly different from the variance/mean ratio before the pulses. The smaller pulse to -70 mV produced a much smaller increase of $[\text{Ca}^{2+}]$ and virtually no inactivation of release (a, b) at this very low level of activation²⁷. Sarcomere spacing, $3.3 \mu\text{m}$. The $\Delta[\text{Ca}^{2+}]$ calibration is 100 nM for a and b, 200 nM for c. The white bar in a represents $10 \mu\text{m}$. $[\text{Ca}^{2+}]$ was estimated^{12,13} from the fluorescence signal F , assuming resting $[\text{Ca}^{2+}] = 80$ nM. $\Delta[\text{Ca}^{2+}]$ was essentially proportional to ΔF for all Figs except for Fig. 1c. Internal solution (mM): 102.5 Cs-glutamate, 5.5 MgCl_2 , 5 Na_2ATP , 4.5 Na-tris-maleate, 13.2 Cs-tris-maleate, 0.1 EGTA, 5 Na_2 -creatine phosphate, 5 glucose, 50 μM fluo-3, pH 7.0. External solution (mM): 125 TEA-methanesulfonate, 5 Cs-HEPES, 2 CaCl_2 , 3×10^{-7} M TTX, pH 7.0; temperature 20°C .

‡ To whom correspondence should be addressed.

FIG. 2 Subcellular characterization of Ca^{2+} -release events in skeletal muscle reveals stochastic recruitment on depolarization. *a*, Localization of the position of the transverse tubules (TT) using 40 μM extracellular sulho-rhodamine B¹⁶ (Polysciences, Rydal, PA) (left, red pseudocolour). The myoplasmic loading of fluo-3 produces a faint striated pattern in fluorescence images with the repeat equal to the sarcomere spacing of 2.9 μm (centre panel, green). The superimposed line-scan images (right, acquired simultaneously using 488-nm and 567-nm illumination from a krypton-argon laser on a Zeiss LSM410 confocal microscope) locate the TT in the centre of the dark regions of the fluo-3 image. (Note that, in Figs 1 and 3, the extremely faint fluo-3 pattern along the scan line in the resting fibre was removed by subtraction). *b*, Test of randomness of Ca^{2+} -release event occurrence during depolarization. Abscissa shows the number (n) of Ca^{2+} -release events and the ordinate shows the number of sarcomeres with n Ca^{2+} -release events during the voltage-activated $[\text{Ca}^{2+}]$ transient. Open circles present data obtained from the fibre shown in Fig. 1; filled circles connected by the line show the discrete Poisson distribution fit to the data (not statistically different from the data, χ^2 test, $P > 0.05$). *c*, Voltage dependence of Ca^{2+} -release events. The line fit to the data (filled circles) shows an exponential increase in the occurrence of Ca^{2+} -release events every 4.1 mV (*b* and *c* from fibre in Fig. 1). *d*, Location of Ca^{2+} -release events within the sarcomere. The relative frequency of Ca^{2+} -release events is plotted against the distance from the TT (identified as in *a*). The solid line represents the location of Ca^{2+} -release events evoked by depolarization (153 events), and the broken line shows the location of 42 spontaneous Ca^{2+} -release events (data from fibre in Fig. 3). *e*, $[\text{Ca}^{2+}]$ dependence of voltage-independent Ca^{2+} -release events. A depolarized fibre segment bathed in internal solution (with 50 μM fluo-3) was cut again to produce small notches extending at most about one-quarter of the fibre diameter into the fibre. The bathing solution was changed to internal solution (50 μM fluo-3) plus 50 μM CaCl_2 (mostly complexed by the 100 μM EGTA in the internal solution), 1 μM nifedipine to fully inactivate the voltage sensors, and 5 mM 2,3-butanedione monoxime to block fibre contraction in elevated $[\text{Ca}^{2+}]$ (ref. 31). The frequency of calcium release events was determined from line-scan images of the fibre taken at a notch (open squares) and at distances of about 60 (open diamonds), 130 (open triangles) and 200 μm (open and filled circles) from the notch. The results are plotted as a function of average $[\text{Ca}^{2+}]$ over each line-scan image assuming $[\text{Ca}^{2+}]$ to be 80 nM at the furthest distance from the notch at the start of the experiment. The filled symbols represent results obtained late in the experiment at a distance furthest from the notch after gentle mechanical agitation of the fibre caused a release of calcium from the SR, after which myoplasmic $[\text{Ca}^{2+}]$ was elevated. This allowed measurements at the highest $[\text{Ca}^{2+}]$ to be obtained at the furthest location from the notch, which had previously exhibited the lowest $[\text{Ca}^{2+}]$ and which now exhibited an increased event frequency despite a decreased average SR Ca^{2+} content. Ca^{2+} -release events are visually identified and spatio-temporally located by positioning a moveable $3.2 \mu\text{m} \times 20 \text{ ms}$ box on a video-displayed line-scan image such that each local $\Delta[\text{Ca}^{2+}]$ was spatially centred in the box and started temporally at the edge of the box. Identification by two independent observers agreed to within 12%.



this fibre (Fig. 2c), or even more steeply in other fibres (Fig. 3 legend). The spatial centre of the local $\Delta[\text{Ca}^{2+}]$ evoked by depolarization corresponded closely with the triad position (Fig. 2d), as expected for the junctional location of the SR calcium-release channels^{5,8}. Although discrete Ca^{2+} -release events could not be unambiguously identified in all fibres, their occurrence was nonetheless suggested by an increase in fluorescence fluctuations during the pulse that was greater than that attributable to increased photon noise¹³ (Fig. 1 legend).

Spontaneous local release events were observed at a holding potential of -90 mV (Fig. 3a,b, arrows) in three out of six fibres,

and were also spatially centred at the triad (Fig. 2d). The events at -90 mV appear to have been initiated by a mechanism independent of the transverse tubule voltage sensor, as they were about 45 times more frequent than expected for activation by voltage sensors in this fibre. Additionally, complete inactivation of the voltage sensor by holding fibres at 0 mV (ref. 4) (Fig. 3c,d), even in the presence of the dihydropyridine nifedipine (1 μM ; Fig. 3e,f), which blocks the voltage sensors², did not suppress the Ca^{2+} -release events (Fig. 3e,f). As is the case with cardiac myocytes, a spontaneous Ca^{2+} -release event could correspond to the opening of a single SR Ca^{2+} -release channel¹², except that the

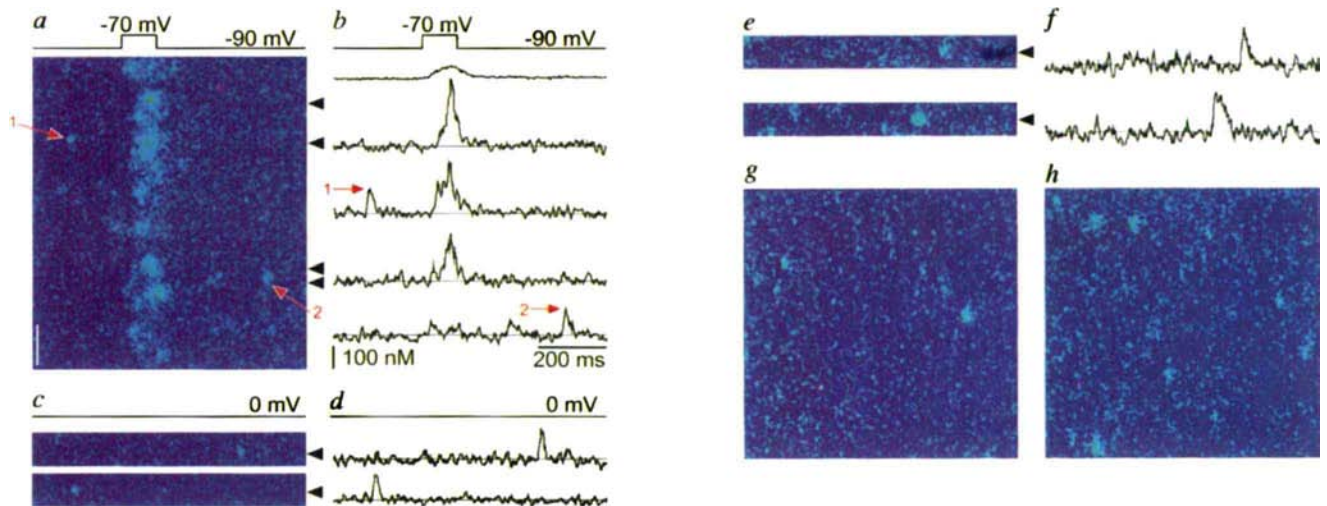
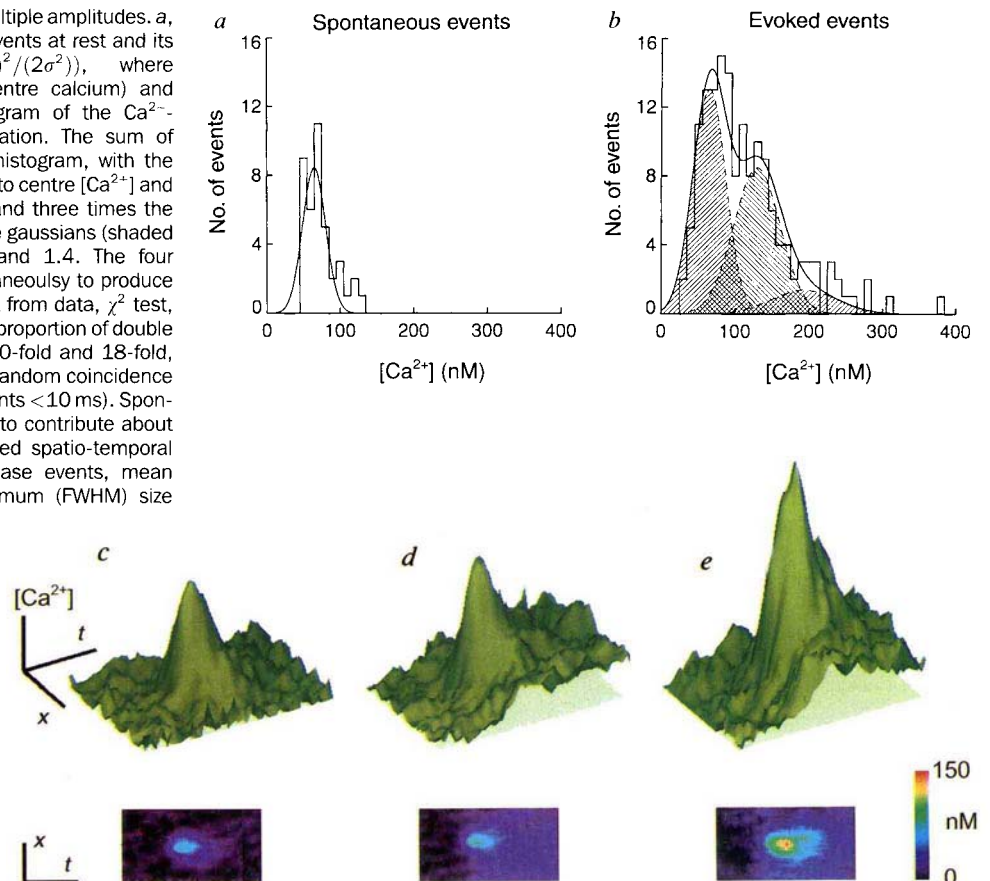


FIG. 3 Dual modes of SR Ca^{2+} -release activation in skeletal muscle. *a, b*, Cut muscle fibres were voltage-clamped at a holding potential of -90 mV and depolarized to -70 mV for 100 ms. *a*, The line-scan image of a single fibre is shown on the left and the timing of the depolarization is indicated by the tracing above the image. *b*, Five traces obtained from the image in *a* are displayed below the voltage protocol (top). The spatially averaged $\Delta[\text{Ca}^{2+}]$ signal and $\Delta[\text{Ca}^{2+}]$ records from four different sarcomeres (marked by arrowheads) are displayed (top to bottom). Spontaneous local SR Ca^{2+} -release events (identified by red arrows in the image and in the second and fourth traces) occurred at a rate of 0.176 per triad per second. On the basis of the voltage dependence of release-event frequency (exponential increase per 2.7 mV depolarization in this fibre) and on the release-event frequency of 6.15 per triad per second observed during the pulses to -70 mV, the voltage-dependent process would be expected to give an event rate of only 0.004 per triad per second at -90 mV, which is about 45

times lower than observed. Sarcomere length $3.2 \mu\text{m}$. *c, d*, Spontaneous local SR Ca^{2+} -release events in fully depolarized fibres at 0 mV holding potential. *c*, Line-scan images of a spontaneous event recorded from the same fibre as in *a* and *b* (top) before fibre polarization and from another fibre (bottom). *d*, Single sarcomere traces showing the spontaneous events in *c*. *e, f*, Nifedipine ($1 \mu\text{M}$) does not suppress spontaneous local SR Ca^{2+} -release events in fully depolarized fibres. *e*, Line-scan images of two spontaneous events recorded from the same fibre in the presence of nifedipine. *f*, Single sarcomere traces showing the spontaneous events in *e*. *g, h*, Caffeine increases the rate of occurrence of spontaneous Ca^{2+} -release events. *g*, Line-scan image of spontaneous Ca^{2+} -release events in the presence of $1 \mu\text{M}$ nifedipine. *h*, Line-scan image of the same fibre as in *g* after adding 0.2 mM caffeine to the nifedipine-containing solution. Note the increase in frequency of events after caffeine addition. White bar in *a*, $10 \mu\text{m}$, and calibration bars apply to all panels. Pseudocolour calibration is the same as Fig. 1.

FIG. 4 Triadic Ca^{2+} -release events exhibit multiple amplitudes. *a*, Amplitude histogram of the Ca^{2+} -release events at rest and its gaussian fit ($y = a \exp(-(|\text{Ca}^{2+}| - b)^2 / (2\sigma^2))$), where $a = 8.43$ (magnitude), $b = 64.71$ nM (centre calcium) and $\sigma^2 = 496$ (variance). *b*, Amplitude histogram of the Ca^{2+} -release events observed during depolarization. The sum of three gaussian profiles was fitted to the histogram, with the first identical to that shown in *a* with respect to centre $[\text{Ca}^{2+}]$ and σ^2 , and the second and third having two and three times the centre $[\text{Ca}^{2+}]$ and σ^2 , respectively. The three gaussians (shaded regions) had magnitudes of 13.1, 8.5 and 1.4. The four gaussians in *a* and *b* were adjusted simultaneously to produce the best overall fit (not statistically different from data, χ^2 test, $P > 0.05$) constrained as noted above. The proportion of double and triple events in *b* was much higher (10-fold and 18-fold, respectively) than expected on the basis of random coincidence (interval between successive poissonian events < 10 ms). Spontaneous Ca^{2+} -release events are expected to contribute about 3% of the total events in *b*. *c-e*, Averaged spatio-temporal pattern of: *c*, 20 spontaneous Ca^{2+} -release events, mean $\Delta[\text{Ca}^{2+}]$ of 62 nM, full-width at half-maximum (FWHM) size $2.37 \mu\text{m}$ and duration 26 ms; *d*, 32 smaller evoked Ca^{2+} -release events, mean $\Delta[\text{Ca}^{2+}]$ of 71 nM FWHM size $2.27 \mu\text{m}$ and duration 32 ms; and *e*, 21 larger evoked Ca^{2+} -release events, mean $\Delta[\text{Ca}^{2+}]$ of 140 nM, FWHM size $2.34 \mu\text{m}$ and duration 34 ms. Note that the $[\text{Ca}^{2+}]$ increases in the regions around the evoked Ca^{2+} -release events (*d* and *e*), owing to many adjacent Ca^{2+} -release events in and out of the plane of focus being evoked by the depolarization, while the area around the spontaneous Ca^{2+} -release events remains unchanged (*c*). Scale bars: *x*, $5 \mu\text{m}$; *t*, 40 ms; $[\text{Ca}^{2+}]$, 30 nM.



amplitude of the local $\Delta[\text{Ca}^{2+}]$ (a Ca^{2+} spark) due to a single event is about threefold smaller in skeletal muscle than in heart.

The frequency of spontaneous voltage-independent Ca^{2+} -release events increases with myoplasmic $[\text{Ca}^{2+}]$ (Fig. 2e), indicating that these events are probably ligand activated^{19,20} by Ca^{2+} by means of calcium-induced calcium release (CICR). Caffeine (0.2–0.5 mM), a pharmacological potentiator of CICR, increased the frequency of spontaneous Ca^{2+} -release events in depolarized fibres (Fig. 3g, h) by a factor of 4.0 ± 0.7 (mean $\pm$ s.e.m.; 3 fibres). Procaine (1–2 mM), an inhibitor of CICR, completely suppressed all Ca^{2+} -release events in depolarized fibres, both in the presence and absence of caffeine (7 fibres). Ryanodine (10 μM) abolished all spontaneous Ca^{2+} -release events within 10–15 min of application (2 fibres), whereas heparin (20 $\mu\text{g ml}^{-1}$; M_r 4K–6K), the inositol trisphosphate (InsP_3) receptor blocker, diffused into the fibre from the internal solution, did not abolish the Ca^{2+} -release events, even after 10–15 min and within 50–100 μm of a cut made in the fibre to facilitate heparin entry (2 fibres).

The amplitude distribution of the spontaneous Ca^{2+} -release events at -90 mV (Fig. 4a) is relatively narrow, consistent with a single class of unitary Ca^{2+} -release events. In contrast, the amplitude distribution of events during the depolarization to -70 mV (Fig. 4b) is broader, consistent with populations of events having amplitudes of 1, 2 or 3 times that of the spontaneous events. Space–time surface plots were constructed by superimposing and averaging spontaneous events at the holding potential (Fig. 4c) for either small (Fig. 4d) or large (Fig. 4e) events during depolarization. The spatio-temporal distribution and amplitude are similar for the spontaneous voltage-independent events and small-voltage-activated events (Fig. 4c, d), suggesting that similar amounts of calcium were released in these events activated by different mechanisms.

The population of smallest-amplitude events during depolarization was similar in amplitude and spatio-temporal characteristics to the spontaneous events, and thus could correspond to the opening of a single SR Ca^{2+} -release channel triggered by the voltage sensor. The population of larger-amplitude Ca^{2+} -release events during depolarization would then correspond to the simultaneous openings of two or more channels. As the frequency of occurrence of double- or triple-amplitude events was 10–18 times larger than expected for random coincidence of single independent events, the double- and triple-amplitude events during depolarization cannot be attributed solely to channels opened by independent voltage sensors. Instead, our results indicate that secondary activation of SR calcium-release channels arises as a consequence of the voltage-gated release process. Although positive feedback of released calcium on the voltage sensors could contribute to this release²¹, CICR activated by high $[\text{Ca}^{2+}]$ near the release channel is a likely factor^{22,23}, as we have shown. A checkerboard-like double row of ryanodine receptors coupled to transverse tubule voltage sensors alternating with ryanodine receptors that are not coupled to voltage sensors²⁴ provides a structural basis for our results. If coupled receptors require voltage activation, and uncoupled receptors are ligand activated, spontaneous events would be of unitary amplitude as a result of the opening of single ligand-gated channels, whereas multiple-amplitude events could occur during depolarization as a result of the opening of both voltage-gated and adjacent ligand-gated channels. Discrete voltage-gated calcium-release events in skeletal fibers have recently been reported²⁵. Our results demonstrate the existence and interaction of both voltage-gated and ligand-gated elementary SR Ca^{2+} -release events in functioning skeletal muscle fibres. □

Received 24 July; accepted 17 November 1995.

- Schneider, M. F. & Chandler, W. K. *Nature* **242**, 244–246 (1973).
- Rios, E. & Brum, G. *Nature* **325**, 717–720 (1987).
- Tanabe, T. et al. *Nature* **328**, 313–318 (1987).
- Chandler, W. K., Rakowski, R. F. & Schneider, M. F. *J. Physiol., Lond.* **254**, 285–316 (1976).
- Franzini-Armstrong, C. & Jorgensen, A. O. *Rev. Physiol.* **56**, 509–534 (1994).
- Tanabe, T., Beam, K. G., Adams, B. A., Niidome, T. & Numa, S. *Nature* **346**, 567–569 (1990).

- Lai, F. A., Erickson, H. P., Rousseau, E., Liu, Q.-Y. & Meissner, G. *Nature* **331**, 315–319 (1988).
- Inui, M., Saito, A. & Fleischer, S. *J. Biol. Chem.* **262**, 1740–1747 (1987).
- Imagawa, T., Smith, J. S., Coronado, R. & Campbell, K. P. *J. Biol. Chem.* **262**, 16636–16643 (1987).
- Takeshima, H. et al. *Nature* **339**, 439–445 (1989).
- Minta, A., Kao, J. P. Y. & Tsien, R. Y. *J. Biol. Chem.* **264**, 8171–8178 (1989).
- Cheng, H., Lederer, W. J. & Cannell, M. B. *Science* **262**, 740–744 (1993).
- Cannell, M. B., Cheng, H. & Lederer, W. J. *Biophys. J.* **67**, 1942–1956 (1994).
- Cannell, M. B., Cheng, H. & Lederer, W. J. *Science* **268**, 1045–1049 (1995).
- Harkins, A. B., Kurebayashi, N. & Baylor, S. M. *Biophys. J.* **65**, 865–881 (1993).
- Endo, M. *J. Physiol., Lond.* **185**, 224–238 (1966).
- Huxley, A. F. & Taylor, R. E. *J. Physiol., Lond.* **144**, 426–441 (1958).
- Maylie, J., Irving, M., Sizto, N. L. & Chandler, W. K. *J. Gen. Physiol.* **89**, 83–141 (1987).
- Fleischer, S. & Inui, M. *Rev. Biophys. Biophys. Chem.* **18**, 333–364 (1989).
- Meissner, G. *Rev. Physiol.* **56**, 485–508 (1994).
- Pizarro, G., Csemernoch, L., Uribe, I., Rodríguez, M. & Rios, E. *J. Gen. Physiol.* **97**, 913–947 (1991).
- Rios, E. & Pizarro, G. *News Physiol. Sci.* **3**, 223–228 (1988).
- Stern, M. D. *Biophys. J.* **63**, 497–517 (1992).
- Block, B. A., Imagawa, T., Campbell, K. P. & Franzini-Armstrong, C. *J. Cell Biol.* **107**, 2587–2600 (1988).
- Tsugorka, A., Rios, E. & Blatter, L. A. *Science* **269**, 1723–1730 (1995).
- Escobar, A., Monck, J., Fernandez, J. & Vergara, J. *Nature* **367**, 739–741 (1994).
- Klein, M. G., Simon, B. J. & Schneider, M. F. *J. Physiol., Lond.* **425**, 599–626 (1990).
- Irving, M., Maylie, J., Sizto, N. L. & Chandler, W. K. *J. Gen. Physiol.* **89**, 1–40 (1987).
- Schneider, M. F. & Simon, B. J. *J. Physiol., Lond.* **405**, 727–745 (1988).
- Simon, B. J., Klein, M. G. & Schneider, M. F. *J. Gen. Physiol.* **97**, 437–471 (1991).
- Maylie, J. & Hui, C. S. *J. Physiol., Lond.* **442**, 551–567 (1991).

ACKNOWLEDGEMENTS. We thank R. Bloch, W. K. Chandler, J. Kao and R. Matteson for comments on the manuscript. This work was supported by the NIH and MDA.

Cloning and characterization of the thyroid iodide transporter

Ge Dai, Orly Levy & Nancy Carrasco*

Department of Molecular Pharmacology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA

IODIDE (I^-) is an essential constituent of the thyroid hormones T_3 and T_4 , and is accumulated by the thyroid. The transport of iodide, the first step in thyroid hormone synthesis, is catalysed by the Na^+/I^- symporter, an intrinsic membrane protein that is crucial for the evaluation, diagnosis and treatment of thyroid disorders. Although several other important thyroid proteins involved in hormone synthesis have been characterized, the Na^+/I^- symporter has not. Here we report the isolation of a complementary DNA clone that encodes this symporter, as a result of functional screening of a cDNA library from a rat thyroid-derived cell line (FRTL-5) in *Xenopus laevis* oocytes. Oocyte microinjection of an RNA transcript made *in vitro* from this cDNA clone elicited a more than 700-fold increase in perchlorate-sensitive Na^+/I^- symport activity over background. To our knowledge, this is the first iodide-transporting molecule to have its cDNA cloned, providing a missing link in the thyroid hormone biosynthetic pathway.

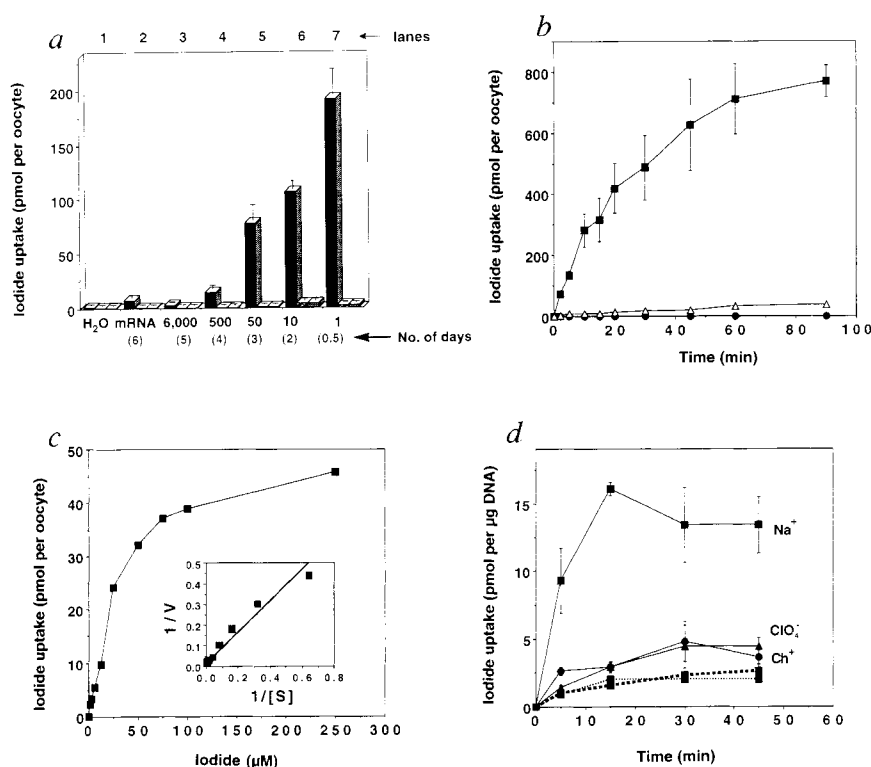
cRNAs synthesized *in vitro* from an FRTL-5 cell line-derived cDNA library were microinjected into oocytes and assayed for perchlorate-sensitive Na^+/I^- symport (NIS) activity. Positive pools containing successively fewer cDNA clones were assayed until a single positive clone was identified (Fig. 1a). The activity elicited by messenger RNA from FRTL-5 cells is also shown (Fig. 1a, lane 2; control (H_2O injected), lane 1). Na^+ dependence was demonstrated by using choline in place of Na^+ ; perchlorate sensitivity was assayed in the presence of both Na^+ and perchlorate. There is a clear correspondence between the decreasing number of clones in the pools, the increase in activity, and the decrease in the lag time before the signal appears after injection.

Iodide accumulation reached nearly 800 pmol per oocyte at saturation (Fig. 1b). Given that the I^- concentration in the

* To whom correspondence should be addressed.

FIG. 1 Expression of NIS in *Xenopus laevis* oocytes and COS cells. **a**, Oocytes were microinjected with water (lane 1), 50 ng mRNA from FRTL-5 cells (lane 2), or ~20 ng cRNA transcripts made *in vitro* from pools containing cDNA clones (numbers indicated under lanes 3–7, without parentheses at bottom of figure). I^- accumulation was assayed in the presence of Na^+ (first bar in set), absence of Na^+ (that is, in the presence of choline, second bar), or in the presence of both Na^+ and perchlorate (third bar) at 45 min as described¹⁷. The number of days that elapsed after microinjection before NIS activity was detectable are indicated. Time course (**b**) and kinetics (**c**) of I^- accumulation were assessed 60 h after oocytes were microinjected with the transcript from the NIS cDNA clone. In **c**, initial velocities of NIS activity were determined at 2 min for various I^- concentrations (1.25–250 μM). Inset: double-reciprocal plot. In **a**, **b** and **c** each bar or experimental point represents the average of >5 oocytes $\pm$ s.e. **d**, Time course of I^- accumulation in COS cells transfected with pSV.SPORT plasmid containing NIS cDNA (continuous lines) or a control insert (dotted line), or with no DNA (dashed line). In **b–d**, I^- accumulation was assayed in the presence of Na^+ (filled squares), absence of Na^+ (and in the presence of choline, filled circles), or in the presence of both Na^+ and perchlorate (filled and open triangles) at the indicated times as described¹⁸. Each point represents the average of triplicate determinations $\pm$ s.e.

METHODS. A directional cDNA library was prepared in the pSPORT (BRL) vector from poly(A)⁺ RNA isolated from FRTL-5 cells. cRNAs were synthesized *in vitro* using T7 RNA polymerase in the presence of m⁷GpppG from pools of cDNA clones. Transcripts were injected into oocytes¹² which were then assayed for perchlorate-sensitive NIS activity as described¹⁷. Clones from one positive pool were subdivided and assayed successively until a single NIS cDNA clone was isolated. COS-7 cells were grown in 6-well plates (37 °C, 5% CO₂) in high-glucose Dulbecco's modified Eagles medium (DMEM) supplemen-



ted with 10% fetal calf serum. Cells were transfected when they reached ~70% density, rinsed with serum-free medium and incubated with Opti-MEM1 (Gibco) containing 0.5 mg ml⁻¹ DEAE-dextran (M_r 2 × 10⁶) and 0.66 μg per well of column-purified (Qiagen) plasmid DNA for 30 min at 37 °C, followed by a 3-h incubation with 0.1 mg ml⁻¹ chloroquine in DMEM. I^- accumulation was assayed 48 h after transfection¹⁷.

transport solution was 50 μM and the functional volume is ~0.5 μl per oocyte, the I^- concentration gradient generated was >30 fold, which is virtually identical to that in the thyroid gland *in vivo*^{1–3}. This signal corresponds to a >700-fold increase in perchlorate-sensitive Na^+/I^- symport activity over background. Na^+/I^- symport rates in oocytes microinjected with the transcript from the NIS clone displayed saturation kinetics (Fig. 1c). The apparent K_m for I^- was 36 μM , a value consistent with the range reported for FRTL-5 cells^{3,4}. COS cells transfected with the NIS cDNA clone exhibited perchlorate-sensitive Na^+/I^- symport activity, in contrast to control cells (Fig. 1d). Thus, the NIS cDNA clone is sufficient to elicit perchlorate-sensitive Na^+/I^- symport activity in both oocytes and mammalian cells.

The nucleotide sequence of the NIS clone (Fig. 2a) indicates that the insert is 2,839 base pairs in length, with a predicted open reading frame of 1,854 nucleotides, a 5'-untranslated region of 109 nucleotides and a 3'-untranslated region of 876 nucleotides. The putative initiation codon ATG contains a purine at position 113 and thus represents a reasonable Kozak consensus sequence⁵. Beginning with methionine at position 1, a long open reading frame codes for a protein of 618 amino acids (relative molecular mass, 65,196). The hydropathic profile and secondary structure predictions^{6–8} of the protein suggest that the cloned cDNA encodes an intrinsic membrane protein with 12 putative transmembrane domains (Fig. 2b). The amino terminus has been placed on the cytoplasmic side, given the absence of a signal sequence, as has the carboxy terminus, which contains a large hydrophilic region of ~70 amino acids. The length of the 12 transmembrane domains in the model ranges from 18 to 28 amino-acid residues¹⁹. Only three charged residues are predicted to lie within transmembrane domains, namely Asp 16, Glu 79 and Arg 208 in helices I, II and VI, respectively¹⁹. By analogy with

other symporters^{9,10}, this suggests that one or more of these residues may play a role in symport activity. Four leucine residues (at positions 199, 206, 213 and 220) appear to comprise a leucine-zipper motif, which may play a part in the oligomerization of subunits in the membrane¹¹. Comparison with other Na^+ -dependent cotransporters shows that NIS shares the highest degree of homology (24.6% amino-acid identity) with the human $Na^+/glucose$ cotransporter¹². A 2.9-kilobase mRNA transcript that hybridizes with the NIS cDNA clone was identified by northern analysis in both FRTL-5 cells and rat thyroid tissue, but not in rat liver, kidney, intestine, brain or heart, suggesting that NIS is primarily expressed in thyroid cells.

The NIS molecule offers a route to diagnosis by thyroid scintigraphic imaging and to the destruction of thyroid cancers with therapeutic doses of I^- isotopes; it is also the conduit that enables large thyroid-cancer-causing doses of I^- isotopes to reach the gland after exposure to nuclear fallout, as in Chernobyl^{13–15}. The characterization of NIS will be invaluable for analysing the molecular basis of radiation-induced thyroid tumorigenesis, of the congenital defect in I^- accumulation¹⁶, and of thyroid autoimmune disorders^{1,2}. □

Received 22 September; accepted 27 November 1995.

- Werner, S. C. & Ingbar, S. in *The Thyroid: A Fundamental and Clinical Text* (eds Braverman, L. E. & Utiger, R. D.) 1–1362 (Lippincott, Philadelphia, 1991).
- DeGroot, L. J. in *Endocrinology* (ed. DeGroot, L. J.) 507–914 (Grune & Stratton, Orlando, FL, 1995).
- Carrasco, N. *Biochim. biophys. Acta* **1154**, 65–82 (1993).
- Weiss, S. J., Philp, N. J. & Grollman, E. F. *Endocrinology* **114**, 1090–1098 (1984).
- Kozak, M. J. *J. Biol. Chem.* **266**, 19867–19870 (1991).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).
- Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).
- Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 222–245 (1974).
- Carrasco, N., Antes, L., Poonian, M. & Kaback, H. R. *Biochemistry* **25**, 4486–4488 (1986).

Several exonic and intronic adenosines in AMPA-receptor subunit GluR-B pre-mRNA are site-selectively deaminated *in vivo* (Fig. 1a). We tested the only molecularly characterized^{8,9} dsRNA-dependent adenosine deaminase, termed dsRAD^{10,11} or DRADA¹², by recombinantly coexpressing it in HEK 293 cells with editing-competent GluR-B minigene B13 (ref. 7). This generated only a slight increase in the editing of the Q/R site (no DRADA, editing $4 \pm 1\%$ ($\pm$ s.d.), $n = 7$; DRADA, $11 \pm 4\%$, $n = 16$), which is edited to $> 99\%$ in GluR-B mRNA and $\sim 85\%$ in pre-mRNA⁷ in the brain. Similarly low Q/R site editing was obtained when GluR-B pre-mRNA synthesized *in vitro* was incubated with nuclear extract of recombinant DRADA-expressing 293 cells (Fig. 1b). However, recombinant DRADA edited the R/G site and the intronic hotspot1 in GluR-B pre-mRNA. This suggested that DRADA may have a role in GluR-B pre-mRNA editing, but that additional dsRNA deaminases may exist

with distinct, perhaps overlapping, substrate specificities.

In an attempt to find a dsRNA adenosine deaminase that is capable of catalysing Q/R-site editing, we screened a rat-brain cDNA library by low-stringency hybridization with a DNA probe for the predicted catalytic domain of DRADA^{8,9}. Identified, cloned cDNAs of more than 3 kb were inserted into a eukaryotic expression vector¹³, the recombinant vectors were co-transfected into HEK 293 cells along with minigene B13, and cell-derived GluR-B-specific reverse transcriptase-polymerase chain reaction (RT-PCR) products were analysed for Q/R-site editing¹⁴. Three vector constructs yielded Q/R site editing at approximately 90%. Co-transfecting minigene B13Δ6 (ref. 7) with a short deletion in the editing-site complementary sequence (ECS) failed to generate Q/R-site edited transcripts, in congruence with the strict dependence of Q/R site editing on this intronic element, both *in vitro*^{7,14-16} and *in vivo*¹⁷. The three recombinant vectors contained

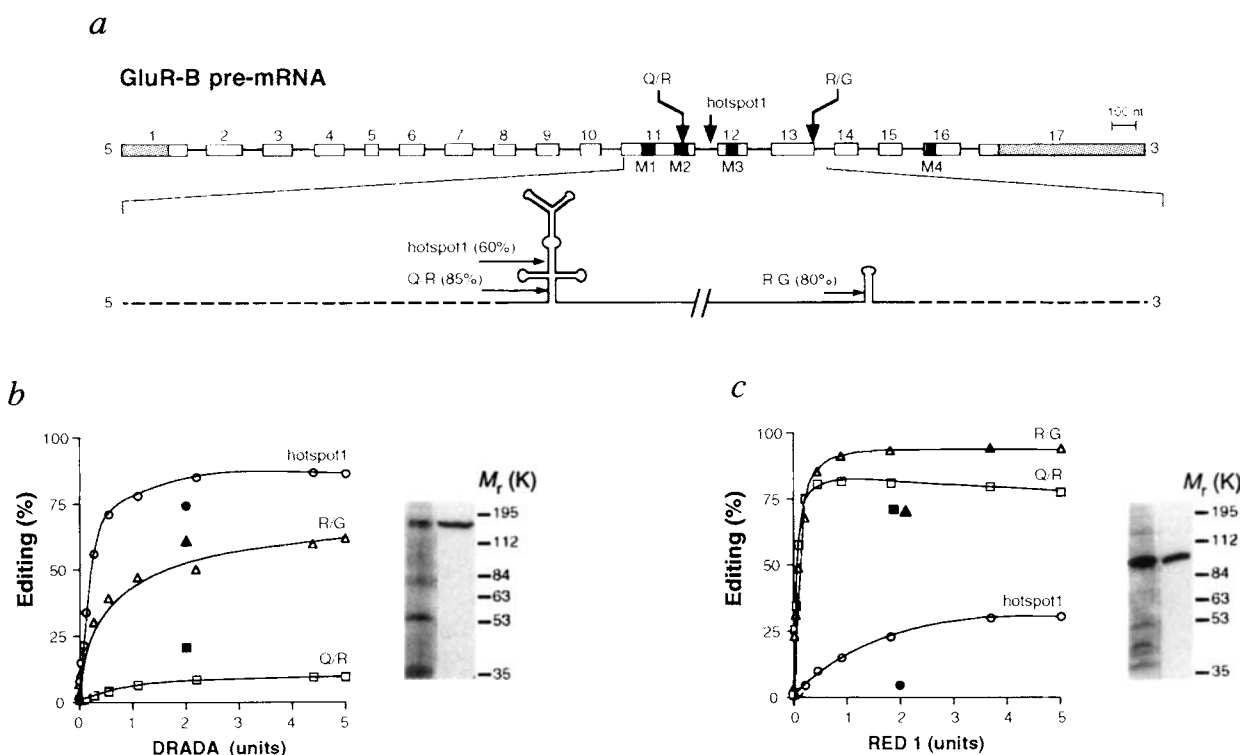


FIG. 1 Activity profiles of recombinant RED1 and DRADA at three editing sites in GluR-B pre-mRNA. **a**, Schematic representation of GluR-B pre-mRNA with the exons numbered and boxed (untranslated regions, grey boxes; scale bar, 100 nucleotides) and introns represented by lines, not drawn to scale²⁸. Exonic sequences encoding subunit segments M1–M4 for membrane insertion are indicated by black boxes. The positions of three editing sites in native GluR-B pre-mRNA are indicated by arrows. The predicted dsRNA structures essential for site-selective RNA editing are depicted below. The extent of editing at the three sites in the pre-mRNA from two-week-old rat brains is listed in parentheses. **b**, **c**, Editing profiles (open squares, Q/R site; open triangles, R/G site; open circles, hotspot1) of the recombinant dsRNA adenosine deaminases DRADA (**b**) and RED1 (**c**) in nuclear extracts of 293 cells. One of three independent experiments, each yielding similar results, is shown. Enzyme units refer to conversion of adenosine to inosine in synthetic, extended dsRNA (1 enzyme unit produces 100 fmol inosine from 4 ng of chloramphenicol acetyltransferase (CAT) dsRNA in 1 h at 30 °C) (ref. 9). The extent of editing indicated by the filled symbols was obtained with two units of purified recombinant deaminase. The gel panels show the single-step immunopurified recombinant deaminases on a silver-stained SDS polyacrylamide gel (left) and by western blot analysis (right).

METHODS. Cloned full-length cDNAs for RED1 (see Fig. 3) and rat DRADA⁹,

both engineered to encode the natural enzymes or enzymes tagged with an N-terminal FLAG epitope and six C-terminal histidine residues, were inserted into a mammalian expression vector¹³. Nuclear extracts²⁰ from HEK 293 cells (ATCC CRL 1573) that transiently expressed recombinant RED1 or DRADA, and the purified recombinant enzymes, were analysed for dsRNA deaminase activity with synthetic CAT dsRNA⁹. *In vitro* editing assays were performed as described previously¹⁴ and cDNAs were amplified by PCR in a two-step amplification procedure with primers *cis55*/PCRK3 (ref. 14) for the first step, and *cis55*/MH50 (MH50, 5'-GACCTGTAGGAAAAATCTAAC-3', Q/R site and hotspot1) or *cis55*/Bint2 (Bint2, 5'-ATCTCTAGACAAACCGT-TAAGAGTC-3', R/G site) for the second amplification. The extent of editing at the different sites was determined by primer extension in gel-purified PCR fragments with primers BRT (Q/R site¹⁴), PEBhot1 (hotspot1, 5'-ATGAA-TATCCACTTGAG-3') and BFF45 (R/G site¹⁴). Signals were quantified by phosphorimager (Fuji, BAS1000). Tagging the enzymes did not alter their activities. For sequence analysis, PCR fragments were cloned and analysed as described previously⁷. Recombinant, 293 cell-expressed deaminases were purified from whole-cell extracts by anti-FLAG affinity chromatography according to the manufacturer (Kodak), and were analysed on 7.5% SDS polyacrylamide gels by silver staining or western blotting (anti-FLAG-M2 antibody, Kodak).

cDNA that encodes a protein with a DRADA-related domain structure (see below), which we termed RED1, for dsRNA-specific editase1. Compatible with the common functional architecture, RED1, like DRADA, converted adenosines to inosine in extended synthetic dsRNA. Nuclear extracts of HEK 293 cells recombinantly expressing these enzymes displayed adenosine deaminase activities that were more than two orders of magnitude greater than extracts from untransfected cells (control extract, 0.006 ± 0.003 units per μg protein, $n = 11$; DRADA, 0.93 ± 0.62 units per μg , $n = 17$; RED1, 2.84 ± 0.49 units per μg , $n = 13$; for unit definition, see Fig. 1 legend). This predicts that the ubiquitous occurrence of deaminase activity on synthetic dsRNA¹⁸, previously attributed to DRADA^{10,11,18}, may reflect the activity of different members¹⁹ of a gene family.

We next compared the activity of the deaminases simultaneously at three editing sites in GluR-B pre-mRNAs synthesized *in vitro*. The extent of editing at these sites was determined by primer extension¹⁴ in a dilution series of nuclear extracts²⁰ from RED1- or DRADA-transfected 293 cells, which had been normalized in activity for adenosine conversion in extended dsRNA⁹. Our results indicated that DRADA produced low-level editing at the GluR-B Q/R site, but edited efficiently at the R/G site and hotspot1 (Fig. 1b). RED1, however, edited both the Q/R and R/G sites efficiently but converted the adenosine in hotspot1 much less efficiently (Fig. 1c). The one-step immunopurified recombinant deaminases displayed largely the same normalized editing activities that were observed in nuclear extracts, the main exception being that purified RED1 did not edit hotspot1. This

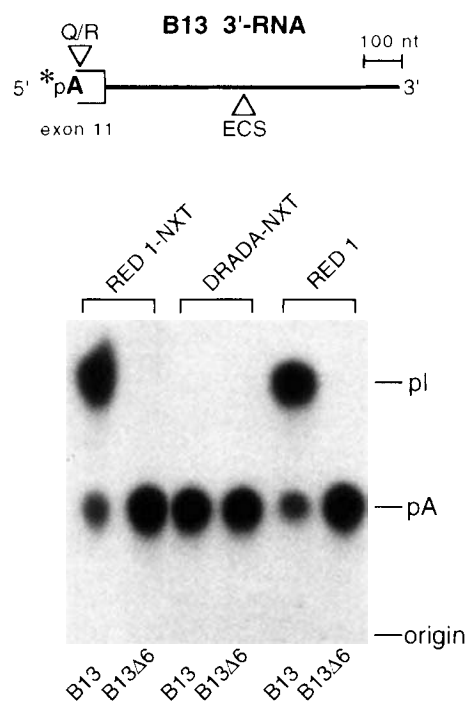


FIG. 2 RED1 deaminates the 5' adenosine of B13 3'-RNA, but DRADA does not. The RNA substrate¹⁶ is depicted on top with the exonic sequence boxed, the intron 11 sequence shown by a line, and the Q/R site and ECS element indicated by arrowheads. The 5' nucleotide is the Q/R site adenosine, and carries a ^{32}P label in its α position (asterisk). Thin-layer chromatography (TLC) shows the edited adenosine in the position of 5'-IMP (pi). NXT, nuclear extract.

METHODS. B13 3'-RNA was synthesized *in vitro* (nucleotides 0–554 of minigene B13; ref. 7) and purified and phosphorylated as described previously¹⁶. Approximately 5 fmol (3×10^3 to 5×10^3 c.p.m.) of labelled RNA was incubated with 5 units of RED1 or DRADA (in nuclear extract from transfected HEK 293 cells or purified, recombinant enzyme) for 3 h at 30°C . RNA was recovered after proteinase K by phenol-chloroform extraction, treated with nuclease P1, and products were resolved by TLC¹⁶.

suggests that cellular factors may affect editing, by interacting with the RNA²¹ or the enzyme.

To demonstrate that RED1 edits the Q/R site by adenosine deamination, we incubated the recombinant enzyme with an *in vitro* transcribed GluR-B pre-mRNA that had the Q/R site adenosine positioned at the 5'-end, and phosphorylated with ^{32}P (ref. 16). Our results (Fig. 2) indicated that RED1 converted this

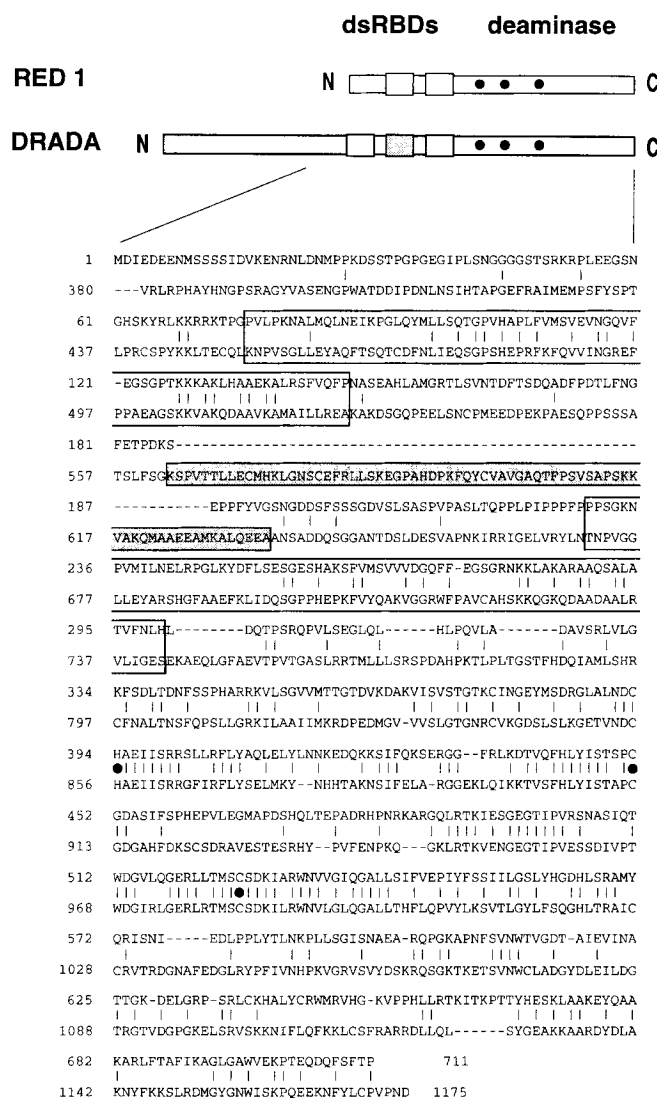


FIG. 3 Domain map and amino acid sequence of RED1 and comparison with DRADA. Domain maps show the location of dsRBDs (boxed) and the putative deaminase domain with Zn^{2+} -chelating residues (filled circles) in the two proteins^{8,9,22}. In the numbered sequence alignment (top, RED1; bottom, DRADA) of the entire RED1 protein and of DRADA starting from residue 380, the dsRBDs are boxed (RED1 dsRBDs 1 and 2 were aligned with DRADA dsRBDs 1 and 3), and gaps are introduced for optimal alignment. The GenBank accession number for RED1 cDNA is U43534. **METHODS.** A rat hippocampal cDNA library constructed in phage λ gt10 was screened with a ^{32}P -labelled DNA probe encoding rat DRADA residues 800–1,030 (ref. 9). Hybridization was in 30% formamide, $5 \times \text{SSC}$ at 37°C , and filters were washed in $0.5 \times \text{SSC}$ at 50°C . The largest DNA inserts of 12 identified recombinant phages were cloned in a mammalian expression vector¹³. The recombinant vectors were co-transfected into HEK 293 cells with GluR-B minigene B13. RT-PCR analysis of GluR-B pre-mRNA sequences from the transfected cells⁷ indicated that three vectors expressed a functional enzyme that edited the Q/R site. The entire coding sequence for RED1 was determined from two of these cDNA clones.

adenosine to inosine with high efficiency ($80 \pm 6\%$, $n = 6$), and the conversion depended on the presence of the ECS element in the RNA substrate. DRADA did not deaminate the adenosine, so the B13 3'-RNA (Fig. 2) allows a clear distinction between the catalytic activities of DRADA and RED1.

As well as determining the extent of adenosine conversion at three editing sites by primer extension (Fig. 1), we assessed the site selectivity of adenosine conversion by the deaminases by sequence analysing 40 cloned RT-PCR products from each of the extract-incubated synthetic RNAs. Analysis of brain-derived GluR-B pre-mRNA sequences has shown that adenosines in the vicinity of the Q/R site (position 0) are also edited^{7,15}. In addition to the positions shown in Fig. 1a, these include the exonic adenosines in positions 4 and 262, 263 (hotspot2, intron 11), edited to $\sim 35\%$ and $\sim 50\%$, respectively^{7,15}. As judged from the analysis of B13 sequences (positions sequenced, -20 to 315), recombinant RED1 (0.7 units) catalysed editing to extents observed in brain at position 4 (35%) and hotspot2 (35%), whereas DRADA (0.7 units) edited mostly the adenosine hotspot1 ($\sim 60\%$) but few other adenosines, and these at low efficiencies ($< 10\%$). Thus the two deaminases leave distinct footprints at and around the Q/R site in GluR-B pre-mRNA, which together might account for the editing pattern

observed in central neurons. With respect to the R/G site and the sequence around it, both deaminases produced similar *in vitro* editing patterns, characterized by high-level conversion of the R/G site adenosine (Fig. 1b) and a lower extent of conversion at position -1 (RED1, 8%; DRADA, 10%), which in the adult rat brain is edited to approximately 10% of the R/G site⁴. Other adenosines in the sequence window inspected (positions -50 to 50 relative to the R/G site) were converted to $< 5\%$. Hence, in contrast to the Q/R site, both DRADA and RED1 are candidate enzymes for catalysing R/G site editing *in vivo*.

The amino acid sequence of RED1 is shown in Fig. 3. DRADA (relative molecular mass, $M_r \sim 130$ K) and RED1 ($M_r \sim 80$ K) share 31% overall sequence identity, largely owing to a conserved carboxy-terminal catalytic domain. This domain contains the short sequence motifs CHAE and PCG, which contain putative Zn^{2+} -chelating residues. These are found in other nucleoside deaminases^{8,9,22} including APOBEC1, which deaminates the cytosine residue in position 6,666 of apolipoprotein B mRNA^{23,24}. Moreover, both DRADA and RED1 carry dsRNA binding domains (dsRBDs)²⁵, the two dsRBDs in RED1 aligning best with the first and third dsRBD of DRADA. DRADA contains a unique, extended amino-terminal domain, the deletion of which (residues 2-392) did not alter the substrate specificity for GluR-B pre-mRNA editing (unpublished results). A database search on RED1 cDNA revealed high sequence identity with the human expressed sequence tags L25485 (ref. 26) and T70335, which appear to originate from human RED1 transcripts.

Northern blot analysis indicated that RED1 mRNA is highly expressed in the brain (Fig. 4a) where it is present in most structures, as judged from *in situ* hybridization (Fig. 4c). These analyses further documented that the expression profile in brain of RED1 mRNA differs from that of GluR-B mRNA²⁷, and that RED1 is also expressed in peripheral tissues (Fig. 4a), suggesting that it is involved in the editing of diverse nuclear transcripts. DRADA gene expression in the brain is more homogeneous⁹, with generally lower transcript levels than RED1 (Fig. 4b), but many neuronal populations appear to coexpress both deaminases.

We conclude that RED1, as judged from both activity profile and brain expression, is an excellent candidate for editing the GluR-B Q/R site in central neurons. We predict that there is a family of dsRNA-dependent adenosine deaminases with distinct yet overlapping expression and substrate profiles, which increase functional protein diversity by recoding nuclear transcripts. □

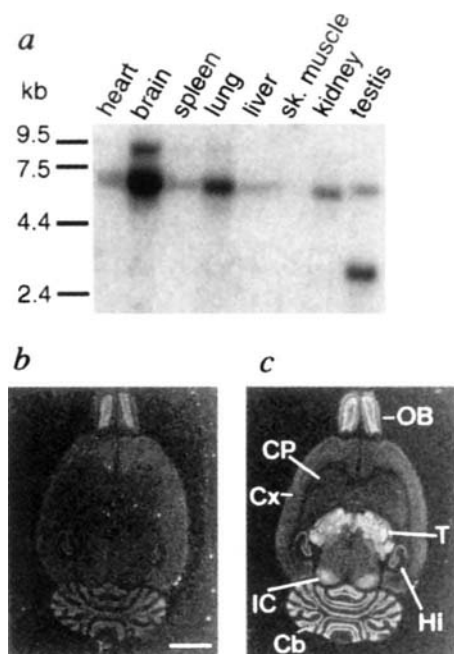


FIG. 4 Expression of RED1 mRNA. a, Northern blot analysis in different rat tissues with size markers indicated on the left. Use of a β -actin probe (not shown) indicated that relative amounts of RNA in lanes differed by no more than twofold. b, c, *In situ* hybridization for DRADA (b) and RED1 mRNA (c) in horizontal brain sections of a 3-week-old rat. Brain structures: Cx, cortex; OB, olfactory bulb; CP, caudate putamen; T, thalamus; HI, hippocampus; IC, inferior colliculus; Cb, cerebellum. Scale bar, 5 mm.

METHODS. A northern blot (Clontech) was hybridized overnight with a ^{32}P -labelled DNA probe (5×10^6 d.p.m. ml^{-1}) encoding RED1 residues 282-515. Hybridization was in 50% formamide, $5 \times$ SSC, 0.1% SDS at $45^\circ C$, and the blot was washed in $0.1 \times$ SSC at $63^\circ C$. Exposure to X-ray film was for 40 h at $-70^\circ C$ with an intensifying screen. *In situ* hybridization²⁹ was performed in horizontal sections of a 3-week-old rat with the 3' end labelled oligonucleotides red1is2 (5'-GCTCGAGCTGTGCGTATAGAAACCTGAGCAGG-GACC-3', complementary to RED1 residues 400-412) and rdrada⁹. Exposure to X-ray film was for 20 days. Control sections hybridized with labelled probe in the presence of a 100-fold excess of the unlabelled oligonucleotide generated no signal.

Received 22 November; accepted 19 December 1995.

1. Scott, J. *Cell* **81**, 833-836 (1995).
2. Sommer, B., Köhler, M., Sprengel, R. & Seeburg, P. H. *Cell* **67**, 11-19 (1991).
3. Köhler, M., Burnashev, N., Sakmann, B. & Seeburg, P. H. *Neuron* **10**, 491-500 (1993).
4. Lomeli, H. et al. *Science* **266**, 1709-1713 (1994).
5. Hollmann, M. & Heinemann, S. A. *Rev. Neurosci.* **17**, 31-108 (1994).
6. Geiger, J. R. P. et al. *Neuron* **15**, 193-204 (1995).
7. Higuchi, M. et al. *Cell* **75**, 1361-1370 (1993).
8. Kim, U., Wang, Y., Sanford, T., Zeng, Y. & Nishikura, K. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11457-11461 (1994).
9. O'Connell, M. A. et al. *Molec. cell. Biol.* **15**, 1389-1397 (1995).
10. Bass, B. L. & Weintraub, H. *Cell* **48**, 607-613 (1987).
11. Bass, B. L. in *The RNA World* (eds Gesteland, R. & Atkins, J.) 383-418 (Cold Spring Harbor Laboratory Press, New York, 1993).
12. Kim, U. & Nishikura, K. *Semin. Cell Biol.* **4**, 285-293 (1993).
13. Schall, T. J. et al. *Cell* **61**, 361-370 (1990).
14. Melcher, T. et al. *J. biol. Chem.* **270**, 8566-8570 (1995).
15. Rueter, S.M., Burns, C. M., Coode, S. A., Mokherjee, P. & Emeson, R. B. *Science* **267**, 1491-1494 (1995).
16. Yang, J.-H., Sklar, P., Axel, R. & Maniatis, T. *Nature* **374**, 77-81 (1995).
17. Brusa, R. et al. *Science* **270**, 1677-1680 (1995).
18. Wagner, R. W. et al. *Molec. cell. Biol.* **10**, 5586-5590 (1990).
19. Patterson, J. B. & Samuel, C. E. *Molec. cell. Biol.* **15**, 5376-5388 (1995).
20. Schreiber, E., Matthias, P., Müller, M. M. & Schaffner, W. *Nucleic Acids Res.* **17**, 6419 (1989).
21. Saccomanno, L. & Bass, B.L. *Molec. cell. Biol.* **14**, 5425-5432 (1994).
22. Bhattacharya, S., Navaratnam, N., Morrison, J. R., Scott, J. & Taylor, W. R. *Trends biochem. Sci.* **19**, 105-106 (1994).
23. Teng, B.-B., Burant, C. F. & Davidson, N. O. *Science* **260**, 1816-1819 (1993).
24. Navaratnam, N. et al. *Cell* **81**, 187-195 (1995).
25. St Johnston, D., Brown, N. H., Gall, J. G. & Jantsch, M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10979-10983 (1992).
26. Cheng, J.-F., Boyartchuk, V. & Zhu, Y. *Genomics* **23**, 75-84 (1994).

27. Keinänen, K. *et al.* *Science* **249**, 556–560 (1990).
 28. Köhler, M., Kornau, H.-C. & Seeburg, P. H. *J. biol. Chem.* **269**, 17367–17370 (1994).
 29. Wisden, W. & Morris, B. J. in *In Situ Hybridization Protocols for the Brain* (eds Wisden, W. & Morris, B. J.) 9–30 (Academic, London, 1994).

ACKNOWLEDGEMENTS. T.M. and S.M. contributed equally to this work. We thank R. Mosbach for database searches; A. Herold and S. Grünewald for DNA sequencing and cell culture; U. Amtmann for *in situ* hybridization; and J. Rami for help with the manuscript. T.M. and S.M. are recipients of a doctoral fellowship from the graduate program of molecular and cellular neurobiology at the University of Heidelberg. This work was supported by grants from SFB317 and HFSP to P.H.S.

A mammalian gene with introns instead of exons generating stable RNA products

Kazimierz T. Tycowski, Mei-Di Shu & Joan A. Steltz*

Department of Molecular Biophysics and Biochemistry,
 Howard Hughes Medical Institute, Yale University School of Medicine,
 295 Congress Avenue, New Haven, Connecticut 06536, USA

THE nucleoli of eukaryotic cells are the sites of ribosomal RNA transcription and processing and of ribosomal subunit assembly. They contain multiple small nucleolar RNAs (snoRNAs), several of which are essential for rRNA maturation¹. The U3, U8 and U13 snoRNA genes are transcribed independently, whereas U14–U24, as well as E3, are located within introns of protein-coding genes, most of whose functions are linked to translation. These snoRNAs are co-transcribed with their host pre-mRNAs and released by processing from excised introns^{1,2}. Here we show that, in addition to U22, seven novel fibrillarin-associated snoRNAs, named U25–U31, are encoded within different introns of the unusually compact mammalian U22 host gene (UHG). All seven RNAs exhibit extensive (12–15 nucleotides) complementarity to different segments of the mature rRNAs, followed by a C/AUGA ('U-turn') sequence. The spliced UHG RNA, although it is associated with polysomes, has little potential for protein coding, is short-lived, and is poorly conserved between human and mouse. Thus, the introns rather than the exons specify the functional products of UHG.

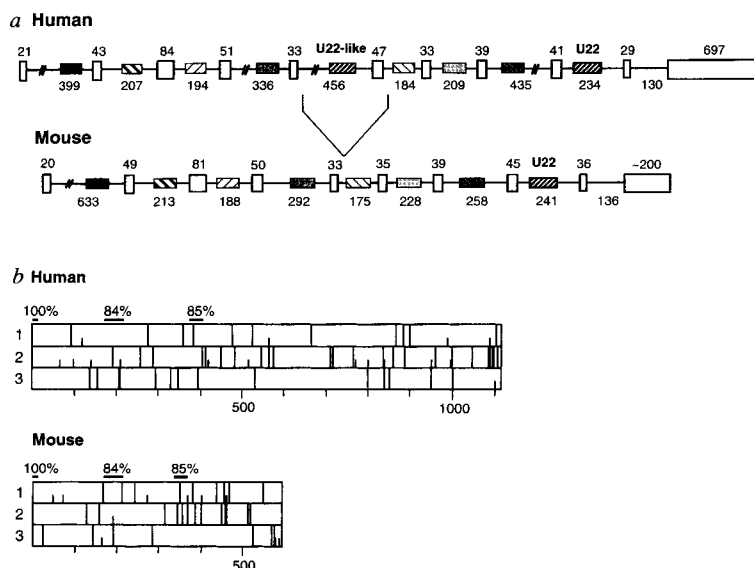
We previously established that nucleolar U22 RNA is essential for the maturation of 18S rRNA³. The human U22 snoRNA gene lies within the penultimate intron of the single-copy gene UHG (for U22 host gene), which encodes a polyadenylated but apparently non-coding RNA. Cloning and partial sequence analysis of genomic and cDNA copies of human UHG revealed multiple unusually short exons and introns³. To search for conserved regions that might provide clues to function, we completed characterization of human UHG and cloned a genomic copy of its mouse homologue (Fig. 1a). The promoters of both genes resemble those of ribosomal protein (rp) genes in that transcription initiates with a C residue within a stretch of pyrimidines embedded in a G + C-rich sequence, suggesting that expression of the UHG and ribosomal protein genes may be co-regulated⁴. Human UHG contains 10 short exons followed by one longer exon. The structure of the mouse gene is similar except that homologues of the human sixth exon and its preceding intron are missing. In humans, this intron contains a U22-like sequence that is 65% identical to U22 snoRNA, but U22-like snoRNA has not been detected by northern blot analyses using oligonucleotide probes (data not shown).

Strikingly, it is the introns rather than the exons of the human and mouse UHGs that exhibit the greatest conservation (Fig. 1). Internal segments of eight out of nine homologous introns are ~90% identical. Conversely, the exons and consequently the spliced UHG RNAs in human (1,114 nucleotides) and mouse (~590 nucleotides) are poorly conserved and have little potential for protein coding (Fig. 1b). The longest open reading frames (ORFs) are 52 and 41 amino acids, respectively, and are both preceded by multiple AUG codons. Three short conserved (>84% identical) sequences found at positions 1–10, 170–215 and 375–403 in the human, and 1–10, 170–216 and 337–363 in the mouse spliced RNAs do not correlate with any of the short ORFs. Yet the human (and presumably the mouse) RNA is polyadenylated and located in the cytoplasm (data not shown).

HeLa cell UHG RNA comigrates in glycerol gradients with small polysomes, from which it can be released by either puromycin or EDTA (data not shown). Interestingly, the level of UHG RNA (normally about 1/20 that of rpL22 mRNA) is increased about 15-fold relative to control cells (Fig. 2, lane 1) within 3 hours of treatment with the protein-synthesis-initiation inhibitor pactamycin (lane 3), or the elongation inhibitors cycloheximide or puromycin (lanes 4 and 5), whereas rpL22 mRNAs and U22

FIG. 1 a, Schematic representation of the human and mouse UHG genes. The open and shaded boxes represent exons and conserved segments of introns corresponding to snoRNAs, respectively. In addition to the exons and introns, about 0.3 kb of the upstream and downstream sequences were determined. The numbers above and below indicate the lengths (in nucleotides) of exons and introns, respectively. Genbank accession numbers for the human and mouse genes are U40580 and U40654, respectively. b, ORFs in three reading frames in the spliced UHG RNAs. Short and long vertical bars within each frame designate translation start and stop codons, respectively. The bars above each schematic diagram indicate the only significantly conserved regions between human and mouse UHG RNAs with the per cent identity shown. Nucleotides are numbered below.

METHODS. Genomic UHG clones were obtained by screening libraries with a human U22 DNA probe³. The 5' end of human UHG RNA was determined by primer extension analysis of HeLa cell poly(A)⁺ RNA using oligonucleotide 5'-CACCAG-TAAGCTCTGTGG-3' as a primer. Before collection, cells were treated with cycloheximide at 20 µg ml⁻¹ for 1.5 h. This analysis revealed that 264 5'-terminal nucleotides of the previously published UHG.3 cDNA³ belong to the first intron. The cap site and the exon/intron composition of mouse UHG have been deduced from comparison of the human and mouse sequences; the end of the mouse UHG RNA was predicted from the presence of AAUAAA and GT-rich sequences.



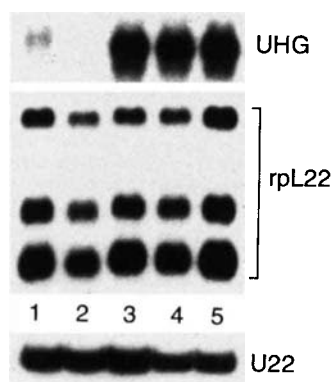
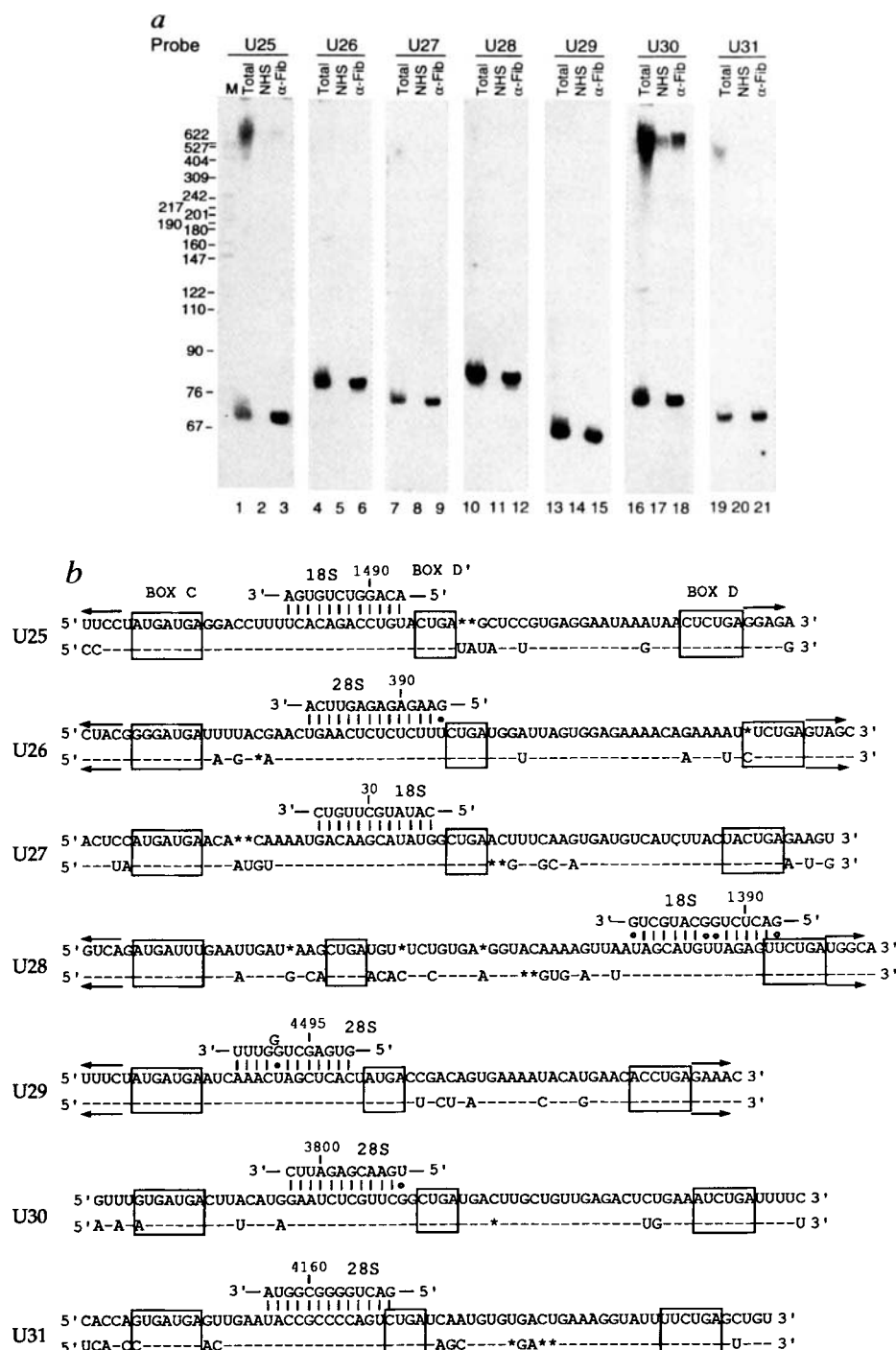


FIG. 2 Inhibition of protein synthesis stabilizes UHG RNA. HeLa cells were grown in the presence of actinomycin D (an RNA-synthesis inhibitor) (lane 2), pactamycin (lane 3), cycloheximide (lane 4) or puromycin (lane 5) at $5 \mu\text{g ml}^{-1}$, 280 ng ml^{-1} , $20 \mu\text{g ml}^{-1}$, or $200 \mu\text{g ml}^{-1}$, respectively, for 3 h before collection. Lane 1 shows control. Poly(A)⁺ RNA was isolated from cytoplasmic extracts¹⁴ and resolved on a 1% agarose/formaldehyde gel. After transferred to a nitrocellulose membrane, the RNA was first hybridized to the UHG cDNA probe and subsequently to the rpL22 cDNA probe¹⁵. Both probes were labelled with [α -³²P]CTP by random priming. snoRNAs were precipitated with anti-fibrillarin (72B9) mAb from nuclear extracts prepared by sonication¹⁶, resolved in an 8% denaturing polyacrylamide gel, and probed for U22 snoRNA by northern blotting.

FIG. 3 Small nucleolar RNAs encoded within UHG introns. *a*, Northern blot analysis of human U25–U31 snoRNAs. HeLa cell nuclear extract prepared by sonication¹⁶ was precipitated with either non-immune human serum (lanes 2, 5, 8, 11, 14, 17, 20) or anti-fibrillarin (72B9) mAb (lanes 3, 6, 9, 12, 15, 18, 21). Precipitated RNAs were fractionated on an 8% denaturing polyacrylamide gel, transferred to a Zeta-Probe membrane and probed with 5'-end-labelled deoxyoligonucleotides complementary to the UHG-encoded snoRNAs indicated at the top. The probes were complementary to nucleotides 22–45, 20–48, 19–44, 34–60, 19–44, 20–46 and 16–42 in U25, U26, U27, U28, U29, U30 and U31, respectively. As a negative control, the membrane that had originally been probed with U31-specific oligonucleotide was stripped and reprobed with oligonucleotides 5'-AATAAGCTCTCATGCCCTCTT-3' and 5'-GTTACAATTACTGGTTC-3', which are complementary to the U22-like sequence in intron 5 (data not shown). None of these probes produced any signal either in total or in α -fibrillarin-precipitated RNA. Lanes 1, 4, 7, 10, 13, 16 and 19 contain total nuclear RNA. DNA size markers (M) in nucleotides are shown on the left. *b*, Sequence comparison of human (upper) and mouse (lower) U25–U31 snoRNAs and their predicted base-pairing interactions with rRNAs. The 5' ends of human U25–U31 were determined by primer extension analyses of HeLa cell nucleolar RNA using primers complementary to nucleotides 22–45, 20–48, 49–67, 34–60, 47–66, 43–62 and 31–50 in U25, U26, U27, U28, U29, U30 and U31, respectively. The 3' ends were inferred from the migration of these RNAs on denaturing polyacrylamide gels and from the location of D boxes, making them accurate to within 3 nucleotides. Identical or missing nucleotides are represented by dashes or asterisks, respectively. Arrows indicate complementary nucleotides that can form terminal stems characteristic for many intron-encoded snoRNAs. Conserved boxes C, D and D' are marked.



snoRNA are essentially unaffected. Thus, association of UHG RNA with polysomes may enable it to be rapidly degraded; turnover of many cellular mRNAs is known to be coupled to translation^{5,6}. Moreover, the densely spaced translational stop codons suggest that the degradation of UHG RNA might be achieved by a nonsense codon-mediated decay mechanism⁶.

To verify that the seven conserved UHG intron segments give rise to small RNAs, we analysed HeLa cell nuclear RNA by northern blotting. Figure 3a shows that, in addition to intron 9 (which encodes U22), the conserved regions of introns 1 to 4 and 6 to 8 are expressed as different small RNAs, all associated with the nucleolar antigen fibrillarin. Fibrillarin binding correlates with the presence in each of these RNAs of conserved sequences called boxes C and D (Fig. 3b) which are found in all metazoan fibrillarin-associated snoRNAs¹. Several of these UHG-encoded RNAs can also form 4-base-pair (bp) terminal stems commonly observed in this class of snoRNAs (Fig. 3b). We designate these novel snoRNAs U25–U31. They range in length from 65 to 76 nucleotides and are similar in abundance to U22 ($\sim 10^4$ molecules per cell). The UHG transcript therefore appears to be a short-lived vehicle for production of the stable U22 (compare lanes 1 and 2 in Fig. 2) and U25–U31 snoRNAs.

U25–U31 all have the potential for extensive base-pairing (12–15 nucleotides) with either 18S or 28S rRNA (Fig. 3b). Similar lengthy interactions have been proposed for the vast majority of the intron-encoded snoRNAs^{7,8}. Even shorter (6–9 bp) interactions between the non-fibrillarin-associated E1, E2 and E3 snoRNAs and rRNAs have been verified by psoralen crosslinking⁹. The U25–U28/rRNA, U30/rRNA, and U31/rRNA base-pairings are each followed by a CUGA sequence in the snoRNA, or AUGA in the U29 case (Fig. 3b). CUGA also appears 3' to 8 out of 11 other complementarities between fibrillarin-associated snoRNAs and rRNA (our unpublished observation and ref. 8) and therefore can serve as an additional predictor of snoRNA/rRNA interactions. As CUGA is an invariant sequence within box D^{1,7}, these snoRNAs possess more than one box D-like sequence. We call the second CUGA box D'.

In 18S and 28S rRNAs, the sites complementary to U25–U31 as well as to other snoRNAs, although scattered, fall within the so-called 'universal cores' of the secondary structures⁷, implying that these snoRNAs may orchestrate the folding of the pre-rRNA for correct processing and ribosomal protein assembly in the nucleolus. The adjacent CUGA constitutes a conserved RNA structural motif, known as the 'U-turn', in the hammerhead ribozyme^{10,11} and transfer RNA¹², and could be used by snoRNAs as a protein-binding signal to facilitate either their interaction or disengagement from rRNA sequences during ribosome biogenesis.

In contrast to most mammalian genes, which discard the bulk of their sequences to produce spliced mRNAs, the eight intron-encoded snoRNAs account for 16 and 20% of the human and mouse UHG transcripts, respectively. Thus, UHG provides a striking example of the utilization of so-called 'junk' DNA, lending support to the existence in eukaryotic organisms of RNA-based gene-expression regulatory systems carried by introns and other apparently non-coding genomic regions¹³. □

13. Mattick, J. S. *Curr. Opin. genet. Dev.* **4**, 823–831 (1994).
14. Geyer, P. K., Meyuhas, O., Perry, R. P. & Johnson, L. F. *Molec. cell. Biol.* **2**, 685–693 (1982).
15. Nucifora, G., Begy, C. R., Erickson, P., Drabkin, H. A. & Rowley, J. D. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7784–7788 (1993).
16. Tyc, K. & Steitz, J. A. *EMBO J.* **8**, 3113–3119 (1989).

ACKNOWLEDGMENTS. We thank Ch. Cheng for Southern blotting of UHG, S. Huang for the mouse genomic library, S. Wolin for advice on polysome analysis, T. McConnell, E. Scharl, W. Tam and other members of the J.A.S. laboratory for discussion, and J. Mermoud, V. Myer, C. Smith and C. Yoo for critical reading of the manuscript. Supported by a grant from the NIH.

Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry

Matthias Wilm*, Andrej Shevchenko*, Tony Houthaeve*, Stephen Breit†, Lothar Schweigerer†‡, Theodore Fotsis†‡ & Matthias Mann*§

*Protein & Peptide Group, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany

†Division of Oncology and Haematology, Children's Hospital, University of Heidelberg, INF 150, 69120 Heidelberg, Germany

MOLECULAR analysis of complex biological structures and processes increasingly requires sensitive methods for protein sequencing. Electrospray mass spectrometry¹ has been applied to the high-sensitivity sequencing of short peptides², but technical difficulties have prevented similar success with gel-isolated proteins. Here we report a simple and robust technique for the sequencing of proteins isolated by polyacrylamide gel electrophoresis, using nano-electrospray^{3,4} tandem mass spectrometry^{5,6}. As little as 5 ng protein starting material on Coomassie- or silver-stained gels can be sequenced. Multiple-sequence stretches of up to 16 amino acids are obtained, which identify the protein unambiguously if already present in databases or provide information to clone the corresponding gene. We have applied this method to the sequencing and cloning of a protein which inhibits the proliferation of capillary endothelial cells *in vitro* and thus may have potential antiangiogenic effects on solid tumours.

The peptide extract produced by *in gel* tryptic cleavage of the protein is passed through a capillary which contains a small volume (about 100 nl) of perfusion sorbent (Fig. 1). The adsorbed peptides are washed extensively and step eluted in a volume of approximately 1 µl directly into the electrospray source for mass spectral acquisition. No electroblotting or chromatographic separation is necessary. The procedure is vastly simplified and seems to be more robust than current Edman or mass-spectrometry sample preparation protocols.

The second key component of this method is the nano-electrospray ion source, a miniaturized electrospray source consisting of a metallized glass capillary needle with a tip of inner diameter 1 µm from which the analyte solution is sprayed. Droplets produced by the nano-electrospray are ~ 100 times smaller in volume than those in conventional electrospray sources, allowing efficient use of the whole sample without loss of material in large droplets from which peptides cannot be ionized³. The ion current is increased even though the flow rate through the capillary needle

Received 18 September; accepted 28 November 1995.

1. Maxwell, E. S. & Fournier, M. J. *A. Rev. Biochem.* **35**, 899–934 (1995).
2. Sollner-Webb, B. *Cell* **75**, 403–405 (1993).
3. Tycowski, K. T., Shu, M.-D. & Steitz, J. A. *Science* **266**, 1558–1561 (1994).
4. Mager, W. H. *Biochem. biophys. Acta* **949**, 1–15 (1988).
5. Greenberg, M. & Belasco, J. G. in *Control of Messenger RNA Stability* (eds Belasco J. & Brawerman G.) 199–218 (Academic, San Diego, 1993).
6. Maquat, L. E. *RNA* **1**, 453–465 (1995).
7. Bachellerie, J.-P. et al. *Trends biochem. Sci.* **20**, 261–264 (1995).
8. Pellizzoni, L., Crosio, C., Campioni, N., Loreni, F. & Pierandrei-Amaldi, P. *Nucleic Acids Res.* **22**, 4607–4613 (1994).
9. Rimoldi, O. J., Raghu, B., Nag, M. K. & Eliceiri, G. L. *Molec. cell. Biol.* **13**, 4382–4390 (1993).
10. Pley, H. W., Flaherty, K. M. & McKay, D. B. *Nature* **372**, 68–74 (1994).
11. Scott, W. G., Finch, J. T. & Klug, A. *Cell* **81**, 991–1002 (1995).
12. Quigley, G. J. & Rich, A. *Science* **194**, 796–806 (1976).

‡ Present addresses: Children's Hospital, University of Marburg, Deutschhausstrasse 12, 35033 Marburg, Germany (L.S.); Laboratory of Biological Chemistry, University of Ioannina Medical School, GR 45110 Ioannina, Greece (T.F.).

§ To whom correspondence should be addressed.

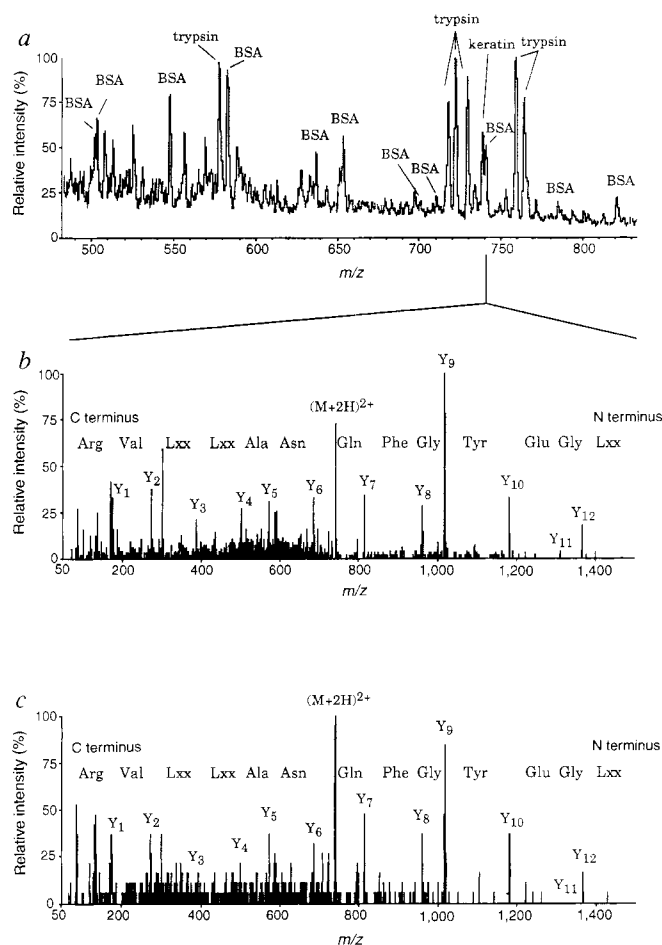


FIG. 1 Testing and sensitivity of the new procedure using BSA. *a*, Peptide ion spectrum of BSA at 800 fmol total material loaded on a gel. Peptide peaks are labelled according to their identification by tandem mass spectrometry. *b*, Fragmentation of the selected doubly charged peptide ion at mass-to-charge m/z 740.5. Fragmentation of tryptic peptides at the amide bonds predominantly produces ion series containing the C terminus (designated Y'_1 , Y'_2 , and so on, see ref. 14). A continuous Y'' ion series¹⁴ could be assigned to the dominant peaks in the spectrum yielding the sequence Lxx-Gly-Glu-Tyr-Gly-Phe-Gln-Asn-Ala-Lxx-Lxx-Val-Arg, where Lxx is either Leu or Ile, which corresponds to a tryptic peptide of BSA. *c*, The same ion as in *b*, obtained in a separate experiment in which 80 fmol BSA was loaded onto a gel that was silver stained.

METHODS. BSA was quantified by amino-acid analysis. Acrylamide gels were prepared using standard protocols and stained with Coomassie blue. *In gel* reduction, acetamidation and tryptic digestion were similar to published procedures^{15,16}. After washing with 100 mM NH_4HCO_3 and acetonitrile, gel pieces were swollen in the digestion buffer containing 50 mM NH_4HCO_3 , 5 mM CaCl_2 and $12.5 \text{ ng } \mu\text{l}^{-1}$ trypsin (Boehringer Mannheim, sequencing grade) at 4°C . After 45 min, the supernatant was aspirated and replaced with 5–10 μl of the same buffer without trypsin to keep gel pieces wet during enzymatic cleavage (37°C , overnight). Peptides were extracted by three changes of 5% formic acid and acetonitrile and dried down. Approximately 100 nl of POROS R2 sorbent (Perseptive Biosystems) was placed in the tip of a pulled GC 100F-10 (CEI, Pangbourne) capillary. Note that the resin is not packed, and that no frit or other micro LC assembling is necessary. A new capillary and a new portion of resin are used for each analysis to avoid cross-contaminations even at the femtomole level. Dried peptide mixture was dissolved in 10 μl 5% formic acid, loaded onto the pre-equilibrated capillary, washed and eluted with 60% methanol in 5% formic acid into the spraying capillary. The elution volume is 10-fold larger than the resin volume, resulting in good peptide recovery. Nano-electrospray was performed on an API III (Perkin-Elmer Sciex, Ontario, Canada) mass spectrometer as described^{4,7}. For precursor ion selection, quadrupole 1 was set to transmit a mass window of 2 Da. Step size for the tandem mass spectra was 0.2 Da, and resolution was set so that fragment masses could be assigned to better than 1 Da.

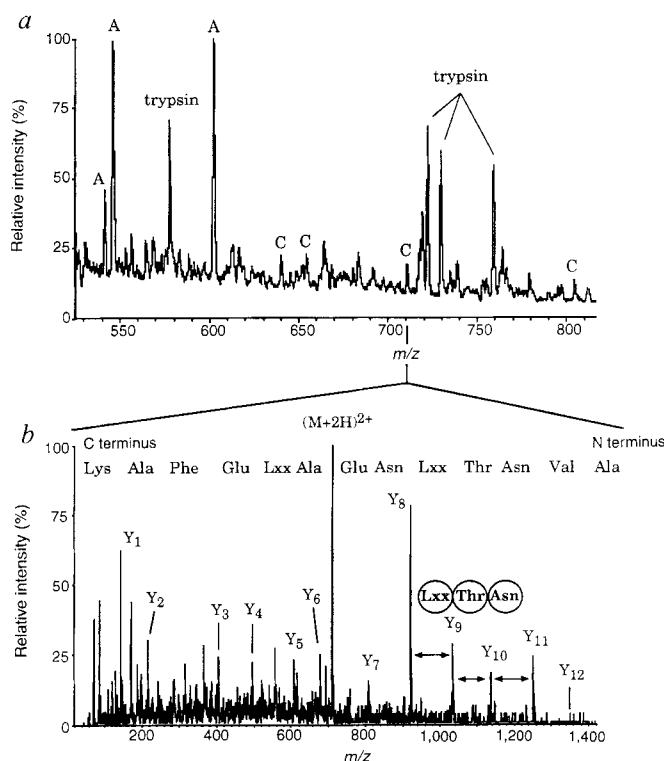


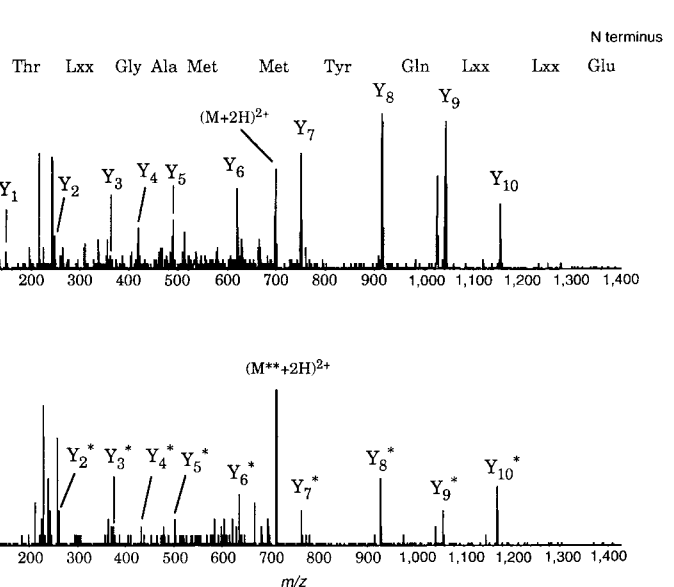
FIG. 2 Identification of protein interaction partners. *a*, Part of the peptide ion spectrum of a tryptic digest of a band of 160K protein excised from polyacrylamide gel. Peptides that were not present in blank were sequenced. The major peaks are tryptic peptides from rabbit antibody light chain (A). *b*, Tandem mass spectrometry of one of the other peptide ions shown in *a*. The masses of the ion and of four fragment ions were assembled by PeptideSearch into a peptide sequence tag⁷. This tag uniquely retrieved carbamoyl-phosphate synthase (M , 164K) from a non redundant, inclusive sequence database (NRDB, maintained by C. Sander, EMBL). As shown in *b*, the complete series of Y'' ions confirms that the peptide sequence found is correct. Mass spectra of three other ions (labelled C in *a*) identified the same protein.

FIG. 3 Sequencing of a peptide from an endothelial cell growth inhibitor. *a*, Tandem mass spectrum of a peptide with m/z 699.3 from the tryptic digest. *b*, Tandem mass spectrum of the same peptide sequenced after esterification⁶ of the total digest mixture. Esterification results in a characteristic mass shift of the C terminus containing fragment ions (Y' ions) by 14 Da and an additional shift of 14 Da for each Asp or Glu residue. The asterisks indicate the incorporated methyl groups. Software-assisted matching of the spectra in *a* and *b* allowed assignment of the sequence with a high degree of certainty.

METHODS. The protein was purified from 20 l conditioned medium from human neuroblastoma cells using the inhibition of proliferation of endothelial cells as an assay. After initial concentration using ultrafiltration, the conditioned medium was subjected to cation-exchanged chromatography and chromatofocusing to enrich the inhibitory activity. Final purification was performed using preparative native PAGE followed by preparative isoelectric focusing in tubes (T.F., in preparation). SDS-PAGE analysis indicated the presence of a protein of M_r 45K (both under reducing and non-reducing conditions) and a yield of approximately 400 ng. Half of the gel piece containing 200 ng of the purified protein was subjected to the new mass-spectrometric sequencing procedure.

is extremely low ($20\text{--}40\text{ nl min}^{-1}$). Very small amounts ($1\text{--}2\text{ }\mu\text{l}$) of the peptide mixture can be subjected to mass spectrometric analysis over 30–120 min. Thus there is sufficient measurement time to optimize tandem mass spectrometric conditions for each of the peptide ions in turn. Tuning the collision energy individually was found to be important to obtain the best possible tandem mass spectra for peptide sequence determination.

We tested the new procedure on a sample of bovine serum albumin (BSA). Figure 1*a* shows the initial peptide ion spectrum, and Fig. 1*b* a tandem mass spectrum, resulting from loading a total of 800 fmol (53 ng) of the protein on a gel. In addition to trypsin and ubiquitous human keratin contaminations, 11 BSA peptides were sequenced. Unambiguous sequence information was also obtained from four peptides when only 80 fmol (5 ng) of BSA was



loaded onto the gel (Fig. 1*c*). In the latter case, the protein band was silver stained because it was hardly visible after Coomassie staining. To the best of our knowledge, there have been no previous reports of protein sequencing from silver-stained gels.

The developed methodology was applied to the identification of protein interaction partners. Approximately 100 ng of an immunoaffinity-purified protein interacting with an RNA binding protein was isolated by gel electrophoresis. The mass spectrum of the peptides derived from this band (Fig. 2*a*) showed high-intensity peaks, which were quickly assigned to trypsin and to antibody (used in the purification step) through database searches by peptide sequence tags⁷. Four other peptides (from Fig. 2*a*) were found to belong to a known protein in the database (Fig. 2*b*). The small amount of sample, the excess of antibody over

TABLE 1 Comparison of the deduced amino-acid sequence to the partial sequences obtained by tandem mass spectrometry

Nucleic acid sequence of cloned gene	Deduced amino-acid sequence	Amino acid sequence obtained by mass spectrometry
GAA CCT GTT CTT TCA CCT GAA CAC AGA	Glu-Pro-Val-Leu-Ser-Pro-Glu-His-Arg	Glu-Pro-Val-Lxx-Ser-Pro-Glu-His-Arg
GAA TTA ATC CAA TAC ATG ATG GCA GGT ATT ACT AAA	Glu-Leu-Ile-Gln-Tyr-Met-Met-Ala-Gly-Ile-Thr-Lys	Glu-Lxx-Lxx-Gln-Tyr-Met-Met-Ala-Gly-Lxx-Thr-Lys
GAC CCA TTT GCA TCA GTT GGT AAT GGT GTT ACA ATT CAC TAC ATG CGT	Asp-Pro-Phe-Ala-Ser-Val-Gly-Asn-Gly-Val-Thr-Ile-His-Tyr-Met-Arg	Asp-Pro-Phe-Ala-Ser-Val-Gly-Asp-Gly-Val-Thr-Lxx-His-Tyr-Met-Arg
ACT CCA TGA TAC TAT GAC CCA GCA ATG AAA	Thr-Pro-Trp-Tyr-Tyr-Asp-Pro-Ala-Met-Lys	Thr-Pro-Trp-Tyr-Tyr-Asp-Pro-Ala-Met-Lys
ACT GAC CTA GAA ACT ATT ACT TTA TTA GCT AAA	Thr-Asp-Leu-Glu-Thr-Ile-Thr-Leu-Leu-Ala-Lys	Thr-Asp-Lxx-Glu-Thr-Lxx-Thr-Lxx-Lxx-Ala-Lys
ATA GTT GCA ATT AAT GTT CCT AAA	Ile-Val-Ala-Ile-Asn-Val-Pro-Lys	Lxx-Val-Ala-Lxx-Asn-Val-Pro-Lys
TGT ATG TCA ATG CCT CTT TCA CGT	Cys-Met-Ser-Met-Pro-Leu-Ser-Arg	Cys-Met-Ser-Met-Pro-Lxx-Ser-Arg

The peptide Cys-Met-Ser-Met-Pro-Lxx-Ser-Arg (Lxx is Leu or Ile), which was at the most C-terminal position (as judged by homology), was used to generate an antisense PCR primer (5'-TGATAAGGCATTGACATAC-3'). As sense primer, a sequence from the 5'-conserved region of the *Mycoplasma arginini* (5'-AGGAATTACGTTTATTTCAG-3') was used to obtain a nearly full-length clone using DNA prepared from the contaminated tumour cells as template. The sequence has been deposited with the EMBL data library (accession no. X93471). The only discrepancy observed was the prediction of Asn from the nucleotide sequence of the cloned protein instead of the mass spectrometrically sequenced Asp in one of the peptides (underlined). Because the identity of Asp is absolutely certain owing to the +14 Da mass shift induced on this amino acid by the derivatization procedure in conjunction with the mass-spectrometric data on the underivatized peptide, it is concluded that deamidation of the amino acid must have occurred during the purification procedure, an event not unusual during protein purification. Thus the amino-acid sequence provided by mass spectrometry was entirely correct.

the protein of interest, and its large size (M_r 164K) would have made identification by conventional techniques exceedingly difficult.

Nano-electrospray mass spectrometry was further used in our search for angiogenesis inhibitors, that is, factors that suppress the formation of new blood vessels^{8,9}. We observed that conditioned media from tumour-cell cultures contaminated by mycoplasma contained a protein that strongly inhibited the *in vitro* proliferation of endothelial cells, the key cell involved in angiogenesis. SDS-polyacrylamide gel electrophoresis (PAGE) analysis after purification revealed a protein with a M_r of 45K. About 200 ng of the protein was subjected to this procedure (Fig. 3). Seven peptides were sequenced in their native and esterified forms, and 73 amino acids were called. A database search of these peptides indicated a high degree of homology to arginine deiminase from *Mycoplasma arginini*. By using the obtained sequence and the homology, we have generated oligonucleotide primers and cloned the complementary DNA by polymerase chain reaction (PCR). A comparison of the deduced amino-acid sequence to the partial sequences obtained by tandem mass spectrometry is

made in Table 1, which shows that all 73 amino acids were sequenced correctly. It is clear from the data that the accuracy and the amount of the amino-acid sequence data would have made cloning by PCR¹⁰ feasible even without homology. Experiments are underway to determine the role of this protein in the inhibition of angiogenesis and its potential in the antiangiogenic treatment of solid tumours.

In conclusion, amino-acid sequence data can now be obtained by mass spectrometry with extremely small quantities of gel-isolated proteins. This procedure is ~10–100 times more sensitive than current techniques, which have a practical limit of 10–50 pmol loaded on the gel for protein microsequencing based on *in gel* or on-blot digestion followed by Edman degradation^{11–13}. It is also much faster and involves fewer purification and manipulation steps. We anticipate that sequencing by nano-electrospray tandem mass spectrometry will thus become a method of choice for protein microcharacterization. In conjunction with genomic sequencing, it will allow the rapid functional characterization of proteins with important biological functions. □

Received 27 September; accepted 28 November 1995.

1. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. *Science* **246**, 64–71 (1989).
2. Hunt, D. F. et al. *Science* **255**, 1261–1263 (1992).
3. Wilm, M. S. & Mann, M. *Int. J. Mass Spectrom. Ion Process.* **136**, 167–180 (1994).
4. Wilm, M. & Mann, M. *Analyt. Chem.* **68**, 1–8 (1996).
5. Biemann, K. & Scoble, H. A. *Science* **237**, 992–998 (1987).
6. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S. & Hauer, C. R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6233–6237 (1986).
7. Mann, M. & Wilm, M. S. *Analyt. Chem.* **66**, 4390–4399 (1994).
8. Fotsis, T. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2690–2694 (1993).
9. Fotsis, T. et al. *Nature* **368**, 237–239 (1994).
10. Walker, J. E. et al. *J. molec. Biol.* **226**, 1051–1072 (1992).
11. Hunkapiller, M. et al. *Nature* **310**, 105–111 (1984).

12. Erdjument-Bromage, H., Lui, M., Sabatini, D. M., Snyder, S. H. & Tempst, P. *Protein Sci.* **3**, 2435–2446 (1994).
13. Matsudaira, P. (ed.) *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Academic, San Diego, CA, 1993).
14. Roepstorff, P. & Fohlman, J. *Biomed. Mass Spectrom.* **11**, 601 (1984).
15. Jenö, P., Mini, T., Hintermann, E. & Horst, M. *Analyt. Biochem.* **224**, 451–455 (1995).
16. Rosenfeld, J., Capdevielle, J., Guillemot, J. C. & Ferrara, P. *Analyt. Biochem.* **203**, 173–179 (1992).

ACKNOWLEDGEMENTS. M.W. and A.S. contributed equally to this work. We thank T. Hyman, E. Izaurralde and A. I. Lamond for cooperation and discussion, and F. Kafatos for criticism and help with the manuscript. This work was supported by the BMBF and constitutes partial fulfillment of the requirement for the Ph.D. of M. Wilm. Work at the University of Heidelberg is supported by grants from Deutsche Forschungsgemeinschaft and Deutsche Krebshilfe.

KNOW YOUR COPY RIGHTS RESPECT OURS

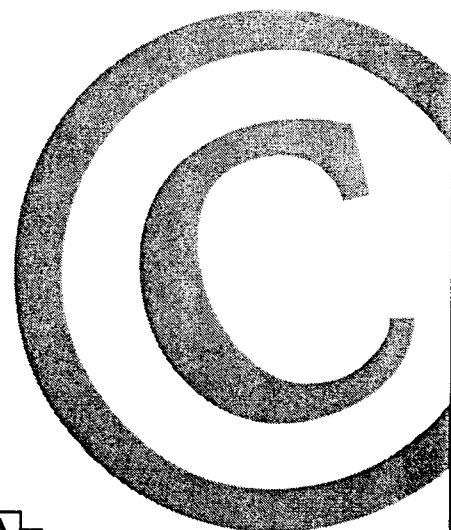
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP

Telephone: 0171 436 5931

Fax: 0171 436 3986

The particulars on peptides

The peptide platter serves up a variety of goodies including antimicrobial peptides, a deamidation detection kit, a molecular-sizing instrument and a new line of separation products for peptides, polypeptides and proteins.

BIOCYTEX-BIOSERVICES offers **custom peptide synthesis**, particularly the synthesis of complex amino-acid sequences. Peptide synthesis can be carried out in either solid-phase or resin (BOC and FMOC strategies), or in the homogeneous phase. In-process control is maintained throughout the synthesis sequence to achieve high purity without altering the peptide sequence. The service includes



Biocytex-Bioservices' custom peptide synthesis.

the supply of lyophilized peptide and a complete amino acid analysis — a mass spectrum is available upon request.

Reader Enquiry No. 100

IntraBiotics Pharmaceuticals, a developer of antimicrobial drugs, offers **new peptide antimicrobials**, called protegrins. Protegrins are peptides derived from the host defence systems of certain mammals; these compounds are capable of killing a wide range of microbial targets through a novel mode of action that is different from that of current antibiotic drugs. Protegrins are lethal to disease organisms, whereas many antibiotics only inhibit microbial growth, but do not kill. These peptide compounds target a broad spectrum of microbes, including Gram-positive bacteria. They are active against multi-drug resistant strains of bacteria because the mode of action differs from those to which the bacteria have become resistant. In addition, researchers could not induce resistance to protegrins in bacteria, in laboratory experiments, whereas the bacteria readily developed resistance to conventional antimicrobials. IntraBiotics expects to develop protegrin-based agents for the treatment of a variety of infectious diseases, including cystic fibrosis, nosoco-

mial infections and gastric ulcers caused by *Helicobacter pylori* bacteria.

Reader Enquiry No. 101

Peptide analysis

Hewlett-Packard offers a new application note entitled *Peptide Mapping Using API-electrospray MS Detection*, which describes how the API-electrospray liquid chromatography/mass spectrometry (LC/MS) system can **map peptides using on-line, mass-spectral detection**. The system provides molecular-weight information that can be used to confirm the primary sequence of a protein by comparing it with a theoretical protein digest, or by searching a database of known proteins. The 12-page note also describes how the system identifies and quantifies peptides that are not fully separated by high-performance liquid

chromatography. The note provides the experimental procedure, instrumentation and instrument conditions, along with the results of the API-electrospray LC/MS and ultraviolet analysis of four protein digests.

Reader Enquiry No. 102

Opioid peptides are important neurotransmitters within the peripheral and central nervous system. They have important roles in the regulation of sensory function (including pain), neuroendocrine activity, the central control of respiration and mood. Chemicon has announced the

availability of **rabbit polyclonal antibodies to opioid receptors**. The antibodies currently available are to the δ -opioid receptor and the μ -opioid receptor. The manufacturer states that these antibodies have been used successfully for immunohistochemistry procedures. Available soon will be an antibody to the κ -opioid receptor. The availability of these antibodies will assist in defining the role of these important molecules.

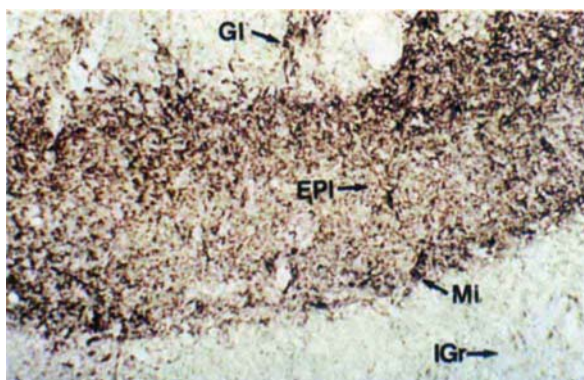
Reader Enquiry No. 103

The new Isoquant **protein deamidation detection kit** from Promega has been developed to quantify the formation of isoaspartic residues in proteins and peptides degraded during long-term storage and handling. The manufacturer states that the sensitivity of this kit is ≤ 2 pmol isoaspartic acid per 50 pmol protein. The kit utilizes the enzyme protein isoaspartyl methyltransferase (PIMT) to catalyse the transfer of a [3 H]methyl group to isoaspartic acid. Spontaneous decomposition of the methylated intermediate releases [3 H]methanol and results in reformation of the succinimide intermediate. Quantification of released [3 H]methanol against a known standard allows a precise determination of the isoaspartic acid content and therefore degree of degradation. The kit has the potential to process up to 30 separate samples in less than 2.5 h.

Reader Enquiry No. 104

A new **temperature-controlled, molecular-sizing instrument** with Dynamics software is now offered by Protein Solutions. The original DynaPro801 molecular sizing instrument is still available for measuring

the size of proteins and other macromolecules in solution. The instrument comes with Windows-based Dynamics software and a temperature-controlled sample chamber. The new design incorporates a thermoelectric cooler, a custom-designed heat exchanger, a semihermetic optics compartment and custom integrated electronics to provide accurate temperature control of the sample cell. The instrument can be used to determine temper-



Antibodies to opioid-peptide receptors from Chemicon.

ature effects and to characterize solution behaviour throughout the temperature range from 4 °C to 40 °C, ± 1 °C. The software provides enhanced data manipulation, including a colour-coding scheme for data interpretation, monomodal and bimodal calculations saved to files and the ability to change solvent viscosity and refractive index.

Reader Enquiry No. 105

Separation

Sartorius has developed a new Sartobind MA5 **membrane adsorber** that purifies and concentrates proteins, peptides and nucleic acids in seconds. This new technology replaces the need for columns and centrifuges and the hours and minutes, respectively, that each of those methods takes to perform the same function. A sample is rapidly loaded onto and eluted from the 3- μ m activated membrane device with a syringe. This membrane adsorber has a capacity of up to 5 mg of protein in volumes from 2 ml to 100 ml. Available chemistries include four types of ion exchanger, IDA for metal chelation and aldehyde for affinity ligands. The manufacturer states that the low price and reproducible operation make this product suitable for sample cleanup (before HPLC or ELISA tests), for rapid isolation of labile proteins from culture media and for screening uncharacterized proteins.

Reader Enquiry No. 106

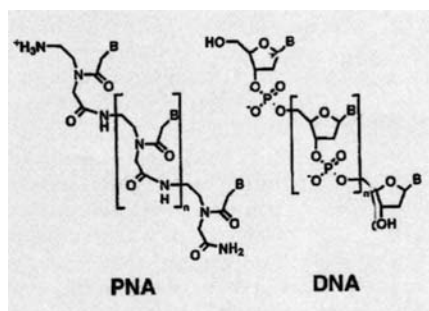
Shandon HPLC has introduced a range of products for the **separation of peptides, polypeptides and proteins**. Based on the company's Hypersil silica technology, the new Hypersil PEP range of columns offers good selectivity, stability and reproducibility. There is a choice of reversed-phase materials, both 100 Å and 300 Å silicas are available in a range of bonded chemistries, as well as in C3, C8 and C18 phases. To enhance peak shape and selectivity the required phase is bonded uniformly and then endcapped

with a short hydrocarbon chain to ensure minimal interaction between the peptide or protein and the silica.

Reader Enquiry No. 107

Miscellaneous

PerSeptive Biosystems has announced the introduction of a new line of **products for the synthesis of peptide nucleic acid (PNA)**. PNA is a new class of informational molecule that has a neutral 'peptide-like' backbone with nucleobases that allow the molecule to hybridize to complementary DNA or RNA with higher affinity and specificity than corresponding oligonucleotides. The unique properties of



A peptide-like backbone with PerSeptive's peptide nucleic acid chemistry.

PNA are said to allow novel molecular biology and biochemistry applications that are unachievable with traditional oligonucleotides and peptides. Present applications include affinity purification, *in situ* hybridization, enhanced PCR, improved Southern blots, single-point mutation analysis and antisense research. The company continues to offer the original form of PNA synthon chemistry. It has developed new PNA monomers with FMOC (9-fluorenylmethoxycarbonyl) and BHOC (benzhydryloxycarbonyl) chemical protecting groups for rapid, efficient solid-phase synthesis using mild chemical conditions and automated nucleic acid synthesis instrumentation. New protocols and soft-

ware modifications for the company's Expedite 8909 nucleic acid synthesis system allow the use of new synthesis chemistry to enable researchers and core facilities to alternate between DNA and PNA synthesis in their labs.

Reader Enquiry No. 108

DLD Diagnostic offers two **ELISA kits** for the quantitative determination of anti-BPI-IgG and semiquantitative determination of anti-BPI-IgA antibodies in serum and plasma. The kits contain 12 $\times$ 8 single, break-apart microwells and colour-coded reagents — assays are said to produce results in about 1.5 h. The new ANCA target, antigen bacterial permeability increasing protein (BPI), is important in ANCA-associated vasculitis, cystic fibrosis, inflammatory bowel diseases with primary sclerosing cholangitis and autoimmune liver disease.

Reader Enquiry No. 109

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

Custom peptide, conjugate and antibody services

PACEMAKER®

When your research depends on high quality at low cost turn to **PACEMAKER®**

All **PACEMAKER®** peptides, whether >70%, >80% or >95% purity, undergo hplc and mass spectrometric analysis.

Peptides are available in quantities from 1mg to 100mg and, with prices from **£115** per peptide and **£90** per conjugate, even the tightest budget is catered for.

Most importantly, you may come to us for advice. Based on many years experience and expertise we will try to find the best approach to fulfilling your custom peptide and antibody requirements.

A F F I N I T I
Research Products Limited

Mamhead Castle, Mamhead
Exeter EX6 8HD, United Kingdom
Tel +44(0) 1626 891010 Fax +44(0) 1626 891090

ADVERTISEMENTS

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT

PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For further information contact Dr. Joe Gray
Phone: +44 191 222 8612
Fax: +44 191 222 8129
Email: JOE.GRAY@newcastle.ac.uk

READER ENQUIRY NO. 12

Gramsch Laboratories

- New: Somatostat. Rec. 2A/B Antisera
- G-Protein Antisera - α , β , γ -subunit
- G-Protein Peptide Fragments
- Opioid Receptor Antisera
- Monoclonal Opioid-Antibody 3E7
- YGGF... - specific; appl.: combinatorial peptide library tests
- Cyclic Nucleotide Antisera
- Antibody/Peptide Production Service

Phone 08138 1769 or 08131 5679-0
Fax 08138 1260 or 08131 5679-24
WWW:ourworld.compuserve.com/homepages/gramsch/gi.htm

Gramsch GmbH
D-85247 SCHWABHAUSEN, Germany

READER ENQUIRY NO. 16

DNA SEQUENCING

£45 per reaction

- PCR products
- M13 inserts
- plasmid inserts

- single strand DNA
- double strand DNA
- no result? no charge!

By research labs for research labs

Molecular Biology Services

King's College School of Medicine and Dentistry
Rayne Institute, 123 Coldharbour Lane, London SE5 9NU
Tel: 0171-346 3126 Fax: 0171-733 3877

READER ENQUIRY NO. 11